MA5800

Feature Guide

Issue	11
Date	2024-06-30

HUAWEI TECHNOLOGIES CO., LTD.

Copyright © Huawei Technologies Co., Ltd. 2024. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.

Trademarks and Permissions

and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.
All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129
People's Republic of China

Website: https://www.huawei.com

Email: support@huawei.com

Security Declaration

Vulnerability

Huawei's regulations on product vulnerability management are subject to the *Vul. Response Process.* For details about this process, visit the following web page:
https://www.huawei.com/en/psirt/vul-response-process
For vulnerability information, enterprise customers can visit the following web page:
https://securitybulletin.huawei.com/enterprise/en/security-advisory

 Copyright © Huawei Technologies Co., Ltd.

About This Document

Introduction to the Document

This document describes access devices' support for the features and the feature implementation modes. This document serves as the reference to network design, network entry tests, and network maintenance.

This document describes a feature from multiple dimensions and helps you learn the following information.

If You Want to Know	See
• What issues does the feature resolve? • What are the feature benefits?	Introduction
According to which protocols or standards the feature is implemented?	Reference Standards and Protocols
• What functions does the feature provide? How are these functions implemented? • Are the feature implementation modes the same as those defined in the protocols? How to implement the scenarios that are not defined or determined in the protocols? • What are the differences between feature implementation modes in different networking modes?	Principles
What scenarios or networking modes does the feature apply to?	Application
How to configure the feature?	Configuration
• What are the maintenance means provided for the feature? • How to locate faults when the feature is abnormal?	Maintenance and Diagnosis
What are the differences between different releases of the feature?	Feature Updates

 Copyright © Huawei Technologies Co., Ltd.

Symbol Conventions

The symbols that may be found in this document are defined as follows.

Symbol	Description
DANGER	Indicates a hazard with a high level of risk which, if not avoided, will result in death or serious injury.
WARNING	Indicates a hazard with a medium level of risk which, if not avoided, could result in death or serious injury.
CAUTION	Indicates a hazard with a low level of risk which, if not avoided, could result in minor or moderate injury.
NOTICE	Indicates a potentially hazardous situation which, if not avoided, could result in equipment damage, data loss, performance deterioration, or unanticipated results. NOTICE is used to address practices not related to personal injury.
NOTE	Supplements the important information in the main text. NOTE is used to address information not related to personal injury, equipment damage, and environment deterioration.

Change History

Issue	Date	Description
11	2024-06-30	Issue with V100R024C00 and later versions.
10	2024-03-31	Issue with V100R023C10 and later versions.
09	2023-09-30	Issue with V100R023C00 and later versions.
08	2022-12-31	Issue with V100R022C10 and later versions. **Added**: Deep energy saving function of optical modules, which is supported by only V100R022C10SPH211.
07	2022-06-30	Issue with V100R022C00 and later versions.
06	2022-03-25	Issue with V100R021C10 and later versions. **Added**: **Feature Specifications and Limitations**
05	2021-07-15	Issue with V100R021C00 and later versions. **Added**: **Access Network Slice**

Copyright © Huawei Technologies Co., Ltd.

Issue	Date	Description
04	2021-03-12	Issue with V100R020C10 and later versions.
03	2020-05-30	Issue with V100R020C00 and later versions. This version and later versions do not support the feature of D-CCAP.
02	2019-09-30	Issue with V100R019C2X and later versions.
01	2019-04-30	This document does not differentiate between R versions starting from V100R019C1X and later versions. The version of this document starts from 01. Updates history of this document due to the updates of the product software, see the Feature Updates section of each chapter.

 Copyright © Huawei Technologies Co., Ltd.

Contents

About This Document iii

1 Feature Specifications and Limitations 1

2 GPON 2

2.1 Introduction to GPON 2
2.2 GPON Working Principle 4
2.3 GPON System Principle 6
2.3.1 GPON Basic Concepts 6
2.3.2 GPON Service Multiplexing 10
2.3.3 GPON Protocol Stacks 11
2.3.4 GPON Frame Structure 13
2.3.5 OMCI 17
2.4 Key GPON Techniques 21
2.4.1 GPON Ranging 21
2.4.2 GPON Burst Optical/Electrical Technology 22
2.4.3 GPON DBA 24
2.4.4 GPON FEC 25
2.4.5 GPON Line Encryption 26
2.4.6 GPON Energy Conservation 27
2.4.7 ONU Automatic Registration and Service Flow Creation 30
2.4.8 ONU Automatic Migration and Aging 32
2.5 GPON Terminal Authentication and Management 33
2.5.1 GPON Terminal Authentication (ONU Is Not Preconfigured) 33
2.5.2 GPON Terminal Authentication (ONU Has Been Pre-configured) 35
2.5.3 GPON Terminal Management 39
2.6 Continuous-Mode ONU Detection 47
2.7 Remote Software Commissioning (PON) 50
2.7.1 Introduction 50
2.7.2 Principles 51
2.7.3 Configuring Remote Software Commissioning (GPON) 53
2.8 GPON Configuration Guide (Profile mode) 54
2.8.1 Configuring a GPON ONT Profile (Profile mode) 54
2.8.1.1 Configuring a DBA Profile 55

 Copyright © Huawei Technologies Co., Ltd.

2.8.1.2 Configuring a GPON ONT Line Profile (Profile Mode) 55
2.8.1.3 Configuring a GPON ONT Service Profile (Profile Mode) 60
2.8.1.4 Configuring a GPON ONT Alarm Profile 67
2.8.2 Configuring a GPON ONT (Profile Mode) 69
2.8.3 Configuring a GPON Port 73
2.9 GPON Configuration Guide (Distributed mode) 76
2.9.1 Configuring a GPON ONT Profile (Distributed Mode) 76
2.9.1.1 Configuring a DBA Profile 77
2.9.1.2 Configuring a GPON ONT Capacity Profile (Distributed Mode) 78
2.9.1.3 Configuring a GPON ONT Alarm Profile 79
2.9.2 Configuring a GPON ONT (Distributed Mode) 81
2.9.3 Configuring a GPON Port 85
2.10 GPON Reference Standards and Protocols 88
3 XG(S)-PON 89
3.1 Introduction to XG(S)-PON 89
3.2 XG(S)-PON Network Structure 90
3.3 XG(S)-PON Working Principle 91
3.3.1 Basic Concepts 91
3.3.2 System Transmit Principles 92
3.3.3 Working Principles of Downstream 93
3.3.4 Working Principle of Upstream 94
3.4 XG(S)-PON Key Technologies 95
3.4.1 Ranging 95
3.4.2 Burst Optical/Electrical Technology 96
3.4.3 DBA 98
3.4.4 FEC 99
3.5 XG(S)-PON Security Technologies 100
3.5.1 Line Encryption 100
3.5.2 ONU Authentication 101
3.5.3 Continuous-Mode ONU Detection 102
3.6 Evolution From GPON to XG(S)-PON 104
3.6.1 Network Evolution Overview 104
3.6.2 Evolution Solution - Using XG(S)-PON Combo 106
3.6.2.1 What Is XG(S)-PON Combo? 106
3.6.2.2 XG(S)-PON Combo Working Principle 107
3.6.2.3 XG(S)-PON Combo Evolution Solution 107
3.6.2.4 XG(S)-PON Combo Configuration Guide 109
3.6.3 Evolution Solution-Using Flex-PON 110
3.6.3.1 What Is Flex-PON? 110
3.6.3.2 What Is 10G PON Port RTU? 111
3.6.3.3 Evolution Solution Overview 115
3.6.3.4 GPON Evolution to XG-PON 118

 Copyright © Huawei Technologies Co., Ltd.

3.6.3.5 GPON Evolution to XGS-PON....119
3.6.3.6 GPON Evolution to XG-PON Combo....120
3.6.3.7 GPON Evolution to XGS-PON Combo....122
3.6.3.8 Flex-PON Configuration Guide....124
3.7 XG(S)-PON Reference Standards and Protocols....125
4 50G-PON....127
4.1 50G-PON Feature Updates....127
4.2 PON Technology Evolution....127
4.3 Introduction to 50G-PON....129
4.4 50G-PON System Architecture....130
4.5 50G-PON Working Principles....131
4.5.1 Basic Concepts....131
4.5.2 System Transmission Principles....132
4.5.3 Downstream Working Principles....132
4.5.4 Upstream Working Principles....133
4.6 Key 50G-PON Technologies....134
4.6.1 Ranging....135
4.6.2 Burst Optical/Electrical Technology....135
4.6.3 DBA....138
4.6.4 FEC....139
4.7 50G-PON Security Policies....140
4.7.1 Line Encryption....140
4.7.2 ONU Authentication....141
4.7.3 Continuous-Mode ONU Detection....143
4.8 50G-PON Evolution Solution....144
4.8.1 50G-PON Network Evolution Overview....144
4.8.2 Working Principles of 50G-PON Combo PON....145
4.8.3 Evolution Solution for XGS-PON Networks....146
4.9 50G-PON Service Configuration....147
4.10 50G-PON Standards and Protocols....148
5 Ethernet Networking Applications....150
5.1 P2P Optical Access....150
5.1.1 What Is P2P Optical Access....150
5.1.2 Networking Application....151
5.1.3 Configuring MDUs Cascaded to an OLT....153
5.2 NE Cascading....155
5.2.1 Introduction to NE Cascading....155
5.2.2 Principle of NE Cascading....156
5.2.3 Configuring NE Cascade and Uplink Transmission....157
5.3 Ethernet OAM....159
5.3.1 Introduction....160
5.3.2 Ethernet OAM Reference Standards and Protocols....161

 Copyright © Huawei Technologies Co., Ltd.

5.3.3 Differences in Implementing Y.1731 and 802.1ag on Access Device 161
5.3.4 CFM (802.1ag and Y.1731) 162
5.3.4.1 CFM Introduction 162
5.3.4.2 CFM Network Application 164
5.3.4.3 CFM Basic Concepts 165
5.3.4.4 CFM Principles 168
5.3.4.4.1 CC Principle 168
5.3.4.4.2 LT Principle 171
5.3.4.4.3 LB Principle 171
5.3.4.4.4 AIS Principles 174
5.3.4.4.5 LLF Principles 175
5.3.4.5 Configuring the Ethernet CFM OAM 179
5.3.5 EFM (802.3ah) 187
5.3.5.1 EFM Introduction 187
5.3.5.2 EFM Basic Concept 189
5.3.5.3 EFM Principle 191
5.3.5.4 EFM Configuration 195
5.3.5.4.1 Configuring an EFM Remote Loopback 195
5.3.5.4.2 Configuring EFM Ethernet Link Monitoring 197
5.3.5.5 EFM Maintenance and Diagnosis 199
5.3.6 PM (Y.1731) 199
5.3.6.1 PM Introduction 199
5.3.6.2 PM Networking Application 201
5.3.6.3 PM Basic Concepts 204
5.3.6.4 PM Principles 207
5.3.6.4.1 Performance Monitoring with Typical VLAN Planning 207
5.3.6.4.2 Performance Monitoring with Typical QoS Planning 211
5.3.6.5 PM Configuration 215

6 MPLS 222

6.1 Overview 222
6.2 MPLS Reference Standards and Protocols 223
6.3 MPLS Overview 225
6.3.1 Introduction 225
6.3.2 Principle 225
6.4 MPLS RSVP-TE 231
6.4.1 Introduction 231
6.4.2 Principle 232
6.5 MPLS OAM 234
6.5.1 Introduction 234
6.5.2 Principle 235
6.6 MPLS TE Reliability 237
6.6.1 RSVP-TE FRR 237

 Copyright © Huawei Technologies Co., Ltd.

6.6.1.1 Introduction 238
6.6.1.2 Principle 238
6.6.2 TE Tunnel Protection Group 245
6.6.2.1 Related Concepts 245
6.6.2.2 Principle 245
6.6.3 CR-LSP Backup 248
6.6.3.1 Related Concepts 248
6.6.3.2 Principle 249
6.7 Configuring the MPLS Service 252
6.7.1 Configuring the Static LSP 252
6.7.2 Configuring the LDP LSP 255
6.7.3 Configure an RSVP-TE LSP 258
6.7.4 Configuring the MPLS RSVP-TE FRR 262
6.7.5 Configuring the MPLS OAM 268
6.7.5.1 Configuration Example for Detection of MPLS OAM for Static LSP Connectivity 269
6.7.5.2 Configuration Example of the MPLS OAM Protection Switching Function 273
7 SRv6 279
7.1 Feature Change History 279
7.2 Introduction 279
7.3 Basic Principles 281
7.4 IS-IS for SRv6 287
7.5 SRv6 VPN 289
7.6 Example for Configuring of SRv6 VPN 292
7.7 Example for Configuring SRv6 EVPN 295
8 VXLAN 298
8.1 VXLAN Overview 298
8.2 Basic VXLAN Concepts 300
8.3 VXLAN Implementation Principles 301
8.4 VXLAN Deployment 303
8.5 Configuring VXLAN 304
8.6 Configuring EVPN-VXLAN 305
9 VPN 310
9.1 PWE3 310
9.1.1 Introduction to PWE3 311
9.1.2 PWE3 Reference Standards and Protocols 311
9.1.3 PWE3 Principle 312
9.1.3.1 Basic Principle of PWE3 312
9.1.3.2 Principle of TDM PWE3 318
9.1.3.3 Principle of ETH PWE3 322
9.1.3.4 Traffic Label Principle 324
9.1.3.5 PW 1:1 Redundancy 326

 Copyright © Huawei Technologies Co., Ltd.

9.1.3.6 PW Redundancy 327
9.1.3.7 PW OAM (VCCV) 328
9.1.4 PWE3 Network Applications 331
9.1.5 Configuring the PWE3 Private Line Service 332
9.1.5.1 Configuring the PWE3 Outer Tunnel 333
9.1.5.2 Configuring the Tunnel Policy 334
9.1.5.3 Configuring the PWE3 Inner PW 335
9.1.5.4 Binding the Service to the PW 339
9.1.5.5 Configuring PW Protection 340
9.1.5.6 Configuring MPLS Tunnel Protection 341
9.1.5.7 Configuring CR-LSP Backup 344
9.2 Native TDM 346
9.2.1 Introduction to Native TDM 346
9.2.2 Native TDM Reference Standards and Protocols 347
9.2.3 Native TDM Principle 347
9.3 VPLS 348
9.3.1 What Is VPLS 349
9.3.2 VPLS Reference Standards and Protocols 349
9.3.3 VPLS Principles 349
9.3.3.1 Introduction to VPLS 349
9.3.3.2 VPLS Layer 2 Functions 356
9.3.3.3 LDP VPLS 359
9.3.3.4 VPLS PW Redundancy 362
9.3.4 VPLS PW Redundancy Applications 364
9.3.4.1 Application of VPLS Individual Access 364
9.3.4.2 Application of VPLS Enterprise Access 366
9.3.4.3 VPLS PW Redundancy for Protecting Multicast Services 367
9.3.4.4 VPLS PW Redundancy for Protecting Unicast Services 371
9.3.5 Configuring VPLS MP2MP Intercommunication 374
9.3.6 Configuration Example of VPLS 378
9.3.6.1 Configuration Example of the VPLS Internet Access Service 379
9.3.6.2 Configuration Example of the VPLS Multicast Service 384
9.3.6.3 Configuration Example of the VPLS Enterprise Private Line Service 389
9.4 VRF 395
9.4.1 Introduction to VRF 395
9.4.2 VRF Principle 396
9.4.3 Configuring IPv4 in VPN 398
9.4.4 Configuring IPv6 in VPN 402
9.5 BGP/MPLS IP VPN 407
9.5.1 Introduction to BGP/MPLS IP VPN 408
9.5.2 BGP/MPLS IP VPN Reference Standards and Protocols 409
9.5.3 Basic BGP/MPLS IP VPN 410

 Copyright © Huawei Technologies Co., Ltd.

9.5.4 Related Concepts for BGP/MPLS IP VPN....411
9.5.5 BGP/MPLS IP VPN Principles....416
9.5.6 Hub & Spoke....418
9.5.7 MCE....422
9.5.8 BGP/MPLS IP VPN Label Distribution Modes....424
9.5.9 BGP Site of Original (SoO)....426
9.5.10 Terms, Acronyms, and Abbreviations for BGP/MPLS IP VPN....427
9.6 GRE....430
9.6.1 Introduction to GRE....430
9.6.2 GRE Basic Principles....431
9.6.3 GRE Principles....435
9.6.3.1 Keepalive Detection....435
9.6.3.2 GRE Security Mechanism....435
9.6.4 GRE Applications....436
9.6.5 GRE Configuration....439
9.6.5.1 Example for Configuring a Static Route for GRE....439
9.6.5.2 Example for Configuring a L3VPN over GRE Tunnel....442
10 Layer 2 Forwarding....445
10.1 MAC Address Management....445
10.1.1 What Is MAC Address Management....445
10.1.2 MAC Address Management Process....447
10.1.3 Smooth MAC Address Expansion....450
10.2 VLAN....452
10.2.1 What Is VLAN....452
10.2.2 VLAN Basic Concepts....452
10.2.3 VLAN Communication Principle....456
10.2.4 VLAN Aggregation (Super VLAN)....458
10.2.5 QinQ VLAN and Stacking VLAN....465
10.2.6 VLAN Translation....468
10.2.7 VLAN Planning Suggestion....470
10.2.8 VLAN Translation Policies Specifications....472
10.2.9 Configuring a VLAN....475
10.2.9.1 Creating a VLAN....477
10.2.9.2 Configuring the VLAN attribute....479
10.2.9.3 Configuring Upstream Port of a VLAN....483
10.2.9.4 Configuring a VLAN Service Profile....483
10.2.9.5 Example: Configuring the QinQ VLAN....488
10.2.9.6 Example: Configuring VLAN Stacking Multi-ISP Wholesale Access....493
10.2.10 VLAN Reference Standards and Protocols....498
10.3 Service Flow....498
10.3.1 Introduction to Service Flow....498
10.3.2 Service Flow Principles....499

 Copyright © Huawei Technologies Co., Ltd.

10.3.3 Configuring Service Flow 504
10.3.3.1 Creating an Ethernet Service Flow 504
10.3.3.2 Creating a GPON Service Flow 507
10.3.3.3 Creating a GPON Service Flow (in Profile Mode with Universal Configurations) 510
10.3.3.4 Creating a GPON Service Flow in Profile Mode with Simplified Configurations 514
10.3.3.5 Configuring Automatic Service Flow Creation (GPON) 518
10.3.4 Maintenance and Diagnosis 527
10.4 Service Port Bundle 527
10.4.1 What Is Service Port Bundle 527
10.4.2 Schematic Diagram For Service Port Bundle 528
10.4.3 Configuring a Service Port Bundle 529
10.5 Layer 2 Forwarding Policy 530
10.5.1 Layer 2 Forwarding Policy Overview 531
10.5.2 Layer 2 Forwarding Policy Principles 531
10.5.3 Configuring a Layer 2 Forwarding Policy 533
10.6 Layer 2 User Bridging 535
10.6.1 Layer 2 User Bridging Overview 535
10.6.2 Layer 2 User Bridging Principles 535
10.6.3 Configuring VLAN-based Layer 2 User Bridging 537
11 Layer 3 Forwarding 539
11.1 IPv4 540
11.1.1 Introduction to IPv4 540
11.1.2 Principles to IPv4 540
11.1.2.1 ICMP 540
11.1.2.2 TCP 542
11.1.2.3 UDP 543
11.1.2.4 RawIP 543
11.1.2.5 Socket 544
11.1.3 IPv4 Reference Standards and Protocols 545
11.2 IPv6 545
11.2.1 Why IPv6 is Required 545
11.2.2 IPv6 network deployment 546
11.2.3 IPv6 Principles 547
11.2.3.1 IPv6 Highlights 547
11.2.3.2 IPv6 Addresses 550
11.2.3.3 IPv6 Packet Format 553
11.2.3.4 ICMPv6 556
11.2.3.5 PMTU 558
11.2.3.6 Dual Protocol Stacks 558
11.2.3.7 TCP6 559
11.2.3.8 UDP6 560
11.2.3.9 RawIP6 560

 Copyright © Huawei Technologies Co., Ltd.

11.2.3.10 IPv6 Neighbor Discovery 561
11.2.4 Configuring Basic IPv6 Information 564
11.2.4.1 Configuring an IPv6 Address for an Interface 566
11.2.4.2 Configuring PMTU 568
11.2.4.3 Configuring IPv6 Neighbor Discovery 569
11.2.5 IPv6 Reference Standards and Protocols 572
11.3 ARP 573
11.3.1 Introduction 574
11.3.2 Principles 575
11.3.2.1 Basic Principles 575
11.3.2.2 Dynamic ARP 581
11.3.2.3 Static ARP 584
11.3.2.4 Gratuitous ARP 586
11.3.2.5 IP Address Conflict Detection 588
11.3.3 Configuring ARP 589
11.3.4 ARP Reference Standards and Protocols 590
11.4 ARP Proxy 591
11.4.1 Introduction to ARP proxy 591
11.4.2 Proxy ARP 591
11.4.3 Configuring ARP Proxy for Interworking 595
11.4.4 ARP Proxy Reference Standards and Protocols 599
11.5 DHCP Server 600
11.5.1 Feature Change History 600
11.5.2 Introduction to the DHCP Server 600
11.5.3 Basic Principles 601
11.5.4 Configuring the DHCP Server 602
11.6 DHCPv6 Server 604
11.6.1 DHCPv6 Server Feature Changes 604
11.6.2 Introduction to DHCPv6 604
11.6.3 DHCPv6 Server Principles 605
11.6.4 Configuring the DHCPv6 Server 607
11.7 DHCP Relay 608
11.7.1 DHCP Relay Overview 608
11.7.2 DHCPv4 Relay Principles 610
11.7.3 DHCPv6 Relay Principle 612
11.7.4 Configuring DHCPv4 Relay 613
11.7.4.1 Configuring the Standard Mode 615
11.7.4.2 Configuring the DHCP Option 60 Mode 618
11.7.4.3 Configuring the DHCP MAC Address Segment Mode 620
11.7.5 Differences Between DHCPv4 and DHCPv6 Configurations 623
11.7.6 DHCP Relay Standards and Protocols Compliance 624
11.8 DHCP Proxy 624

 Copyright © Huawei Technologies Co., Ltd.

11.8.1 What Is DHCP Proxy 624
11.8.2 DHCP Proxy Principles 625
11.8.3 DHCP Proxy Standards and Protocols Compliance 628
11.9 BFD 628
11.9.1 Introduction to BFD 628
11.9.2 Principles to BFD 629
11.9.3 BFD Echo 632
11.9.4 Application Environment 634
11.9.4.1 BFD for IP 635
11.9.4.2 BFD for OSPF 635
11.9.4.3 BFD for IS-IS 636
11.9.4.4 BFD for BGP 637
11.9.4.5 BFD for RSVP 638
11.9.4.6 BFD for LAG 639
11.9.4.7 BFD for PW 639
11.9.5 Configuring the BFD 642
11.9.5.1 Configuring BFD Sessions 642
11.9.5.2 Configuring One-Arm BFD Echo Sessions 645
11.9.5.3 Configuring BFD for IPv4 Static Routes 647
11.9.5.4 Configuring BFD for OSPF 648
11.9.5.5 Configuring BFD for IS-IS 649
11.9.5.6 Configuring BFD for BGP or BGP4+ 651
11.9.5.7 Configuring BFD for RSVP 653
11.9.5.8 Configuring BFD for LAG 655
11.9.6 BFD Reference Standards and Protocols 655
11.10 TWAMP 656
11.10.1 TWAMP Introduction 657
11.10.2 TWAMP Basic Concepts 658
11.10.3 TWAMP Principles 660
11.10.4 TWAMP for ONT 665
11.10.5 Configuring TWAMP 666
11.10.5.1 Configuring TWAMP Responder 666
11.10.5.2 Configuring TWAMP for ONT 667
11.10.6 TWAMP Reference Standards and Protocols 669
11.11 TWAMP Light 669
11.11.1 TWAMP Light Introduction 670
11.11.2 TWAMP Light Basic Concepts 670
11.11.3 TWAMP Light Principles 672
11.11.4 Configuring TWAMP Light Responder 675
11.11.5 TWAMP Light Reference Standards and Protocols 676
11.12 Routing 676
11.12.1 IP Routing Overview 676

 Copyright © Huawei Technologies Co., Ltd.

11.12.1.1 Introduction to Routing 676
11.12.1.2 Routers 677
11.12.1.3 Routing Table and FIB Table 678
11.12.1.4 Routing Protocols 681
11.12.1.5 Priority-based Route Convergence 684
11.12.1.6 Load Balancing and Route Backup 686
11.12.1.7 Re-advertisement of Routing Information 687
11.12.1.8 Default Route 688
11.12.1.9 IP FRR 688
11.12.1.10 ECMP 689
11.12.1.10.1 Introduction to ECMP 689
11.12.1.10.2 ECMP Principle 690
11.12.1.10.3 ECMP Fast Switching 690
11.12.2 Static Route 691
11.12.2.1 Introduction 692
11.12.2.2 Components 692
11.12.2.3 Applications 693
11.12.2.4 Configuration Example of the IPv4 Static Route 695
11.12.2.5 Configuration Example of the IPv6 Static Route 696
11.12.2.6 Example for Configuring IP FRR for an IPv4 Static Route 698
11.12.2.7 Example for Configuring IP FRR for an IPv6 Static Route 700
11.12.3 RIP 702
11.12.3.1 Introduction 703
11.12.3.2 Principles 703
11.12.3.2.1 RIP-1 703
11.12.3.2.2 RIP-2 704
11.12.3.2.3 Timers 704
11.12.3.2.4 Split Horizon 705
11.12.3.2.5 Poison Reverse 707
11.12.3.2.6 Triggered Update 707
11.12.3.2.7 Route Summarization 708
11.12.3.2.8 Multi-Process 709
11.12.3.2.9 RIP NSR 709
11.12.3.2.10 RIP Authentication 709
11.12.3.3 Configuration Example of RIP 710
11.12.3.4 RIP Reference Standards and Protocols 713
11.12.4 RIPng 713
11.12.4.1 Introduction 714
11.12.4.2 RIPng Packet Format 714
11.12.4.3 Timers 716
11.12.4.4 Split Horizon 716
11.12.4.5 Poison Reverse 717

 Copyright © Huawei Technologies Co., Ltd.

11.12.4.6 Triggered Update 717
11.12.4.7 Route Summarization 718
11.12.4.8 Multi-Process and Multi-Instance 719
11.12.4.9 Hot Standby 719
11.12.4.10 IPSec Authentication 719
11.12.4.11 RIPng NSR 720
11.12.4.12 Configuration Example of RIPng 721
11.12.4.13 RIPng Reference Standards and Protocols 725
11.12.5 IS-IS 725
11.12.5.1 Introduction 725
11.12.5.2 Principles 725
11.12.5.2.1 Basic Concepts of IS-IS 726
11.12.5.2.2 Basic Protocols of IS-IS 729
11.12.5.2.3 IS-IS Routing Information Control 736
11.12.5.2.4 IS-IS Overload 739
11.12.5.2.5 IS-IS Fast Convergence 739
11.12.5.2.6 IS-IS LSP Fragment Extension 741
11.12.5.2.7 IS-IS 3-Way Handshake 745
11.12.5.2.8 IS-IS NSR 745
11.12.5.2.9 IS-IS Wide Metric 746
11.12.5.2.10 IS-IS Authentication 747
11.12.5.2.11 IS-IS Control Messages 749
11.12.5.2.12 IS-IS for IPv6 755
11.12.5.2.13 IS-IS Auto FRR 755
11.12.5.3 Configuration Example of IS-IS 756
11.12.5.4 Configuration Example of IS-IS for IPv6 758
11.12.5.5 Example for Configuring IS-IS Auto FRR 760
11.12.5.6 Example for Configuring IPv6 IS-IS Auto FRR 763
11.12.5.7 IS-IS Reference Standards and Protocols 765
11.12.6 OSPF 767
11.12.6.1 Introduction to OSPF 767
11.12.6.2 Principles 769
11.12.6.2.1 Basic Concepts of OSPF 769
11.12.6.2.2 Basic Principles of OSPF 778
11.12.6.2.3 OSPF Route Control 785
11.12.6.2.4 OSPF NSSA 787
11.12.6.2.5 OSPF Fast Convergence 788
11.12.6.2.6 OSPF NSR 790
11.12.6.2.7 OSPF Authentication 790
11.12.6.2.8 OSPF Packet Format 792
11.12.6.2.9 OSPF LSA Format 799
11.12.6.2.10 OSPF IP FRR 805

 Copyright © Huawei Technologies Co., Ltd.

11.12.6.3 Configuration Example of OSPF....807
11.12.6.4 Example for Configuring OSPF IP FRR....809
11.12.6.5 OSPF Reference Standards and Protocols....811
11.12.7 OSPFv3....812
11.12.7.1 Introduction....812
11.12.7.2 Principles....812
11.12.7.2.1 OSPFv3 Fundamentals....812
11.12.7.2.2 Comparison Between OSPFv3 and OSPFv2....818
11.12.7.2.3 OSPFv3 GR....820
11.12.7.2.4 OSPFv3 Authentication....821
11.12.7.2.5 OSPFv3 Packet Format....822
11.12.7.2.6 OSPFv3 LSA Format....828
11.12.7.2.7 OSPFv3 Auto FRR....838
11.12.7.3 Configuration Example of OSPFv3....839
11.12.7.4 Example for Configuring OSPFv3 Auto FRR....843
11.12.7.5 OSPFv3 Reference Standards and Protocols....846
11.12.8 BGP....847
11.12.8.1 Introduction....847
11.12.8.2 Principles....849
11.12.8.2.1 Basic Principle....849
11.12.8.2.2 BGP Route Processing....853
11.12.8.2.3 Peer Group....857
11.12.8.2.4 Route Dampening....858
11.12.8.2.5 Community Attribute....859
11.12.8.2.6 Route Reflector....860
11.12.8.2.7 BGP Confederation....864
11.12.8.2.8 MP-BGP....865
11.12.8.2.9 BGP GR....866
11.12.8.2.10 BGP Dynamic Update Peer-Groups....868
11.12.8.2.11 BGP NSR....870
11.12.8.2.12 4-Byte AS Number....872
11.12.8.2.13 BGP Best-external....876
11.12.8.2.14 BGP ADD-PATH....879
11.12.8.2.15 IP FRR....880
11.12.8.3 Configuration Example of BGP....882
11.12.8.4 Configuration Example of BGP4+....883
11.12.8.5 BGP Reference Standards and Protocols....888
11.12.9 Routing Policy....891
11.12.9.1 Introduction to Routing Policies....891
11.12.9.2 Principles....892
11.12.9.3 Applications....897
11.12.9.4 Configuration Example of the Routing Policy....898

 Copyright © Huawei Technologies Co., Ltd.

11.12.9.5 Routing Policy Reference Standards and Protocols 900
12 Internet Access 901
12.1 DHCP Emulation 901
12.1.1 DHCP Emulation Overview 901
12.1.2 DHCP Emulation Principles 902
12.1.3 DHCP Emulation Usage Scenario 904
12.1.4 Configuring DHCP Emulation 907
12.1.5 DHCP Emulation Reference Standards and Protocols 910
12.2 PPPoE Dialup Emulation 910
12.2.1 PPPoE Dialup Emulation Introduction 910
12.2.2 PPPoE Dialup Emulation Principles 911
12.2.3 PPPoE Dialup Emulation Usage Scenario 913
12.2.4 Configuring PPPoE Dialup Emulation 915
12.2.5 PPPoE Dialup Emulation Reference Standards and Protocols 918
12.3 Traffic Burst Detection 918
12.3.1 Traffic Burst Detection Overview 918
12.3.2 Traffic Burst Detection Principle 918
12.3.3 Traffic Burst Detection Application Scenario 920
12.3.4 Configure Traffic Burst Detection 921
13 Multicast 924
13.1 Introduction to Multicast 924
13.2 Basic Multicast Concepts 926
13.3 Multicast Model 932
13.4 Implementation Principles of Multicast 936
13.4.1 IGMP 936
13.4.1.1 IGMP Protocol 936
13.4.1.2 IGMP Mode 943
13.4.1.3 IGMP Framework on the Device 945
13.4.2 PIM 946
13.4.2.1 Introduction 946
13.4.2.2 PIM-SM 949
13.4.2.3 PIM-SSM 954
13.4.2.4 PIM Control Messages 957
13.4.3 Multicast Forwarding 965
13.4.3.1 Join Flow 965
13.4.3.2 Leave Flow 966
13.4.3.3 Forwarding Framework on the Device 972
13.4.3.4 GPON Multicast Duplication 975
13.4.4 Multicast Upstream Interoperation 977
13.4.4.1 Multicast Cascading 977
13.4.4.2 Ring Network of Uplink Ports 983
13.4.4.3 Dual-homing of Upstream Ports 987

 Copyright © Huawei Technologies Co., Ltd.

13.4.5 Advanced Multicast Technologies....987
13.4.5.1 Multicast Program Management....987
13.4.5.2 Multicast Rights Management....989
13.4.5.3 Multicast CAC....993
13.4.5.4 Multicast QoS....995
13.4.5.5 GPON ONT Multicast....996
13.4.5.6 Multi-instance Multicast....997
13.4.5.7 Charging Mode....1000
13.4.5.8 Double-VLAN Tag Multicast....1004
13.4.5.9 Transparent Transmission of Multicast Data....1006
13.5 IPv6 Multicast....1008
13.5.1 Introduction to IPv6 Multicast....1008
13.5.2 Principle....1008
13.6 Configuring the Multicast Service....1011
13.6.1 Differences Between IPv4 and IPv6 Multicast Configurations....1011
13.6.2 Configuring the Multicast Service on a Single NE....1013
13.6.2.1 Configuring Multicast Global Parameters....1015
13.6.2.2 Configuring the Multicast VLAN and the Multicast Program....1018
13.6.2.3 Configuring the Multicast GPON ONT....1026
13.6.2.4 Configuring a Multicast User....1028
13.6.2.5 (Optional) Configuring the Multicast Bandwidth....1030
13.6.2.6 (Optional) Configuring Multicast Preview....1032
13.6.2.7 (Optional) Configuring Program Prejoin....1034
13.6.2.8 (Optional) Configuring the Multicast Logging Function....1035
13.6.2.9 (Optional) Configuring the Maximum Number of Programs That Can Be Watched by the Multicast User....1039
13.6.2.10 (Optional) Configuring the Maximum Rate for Sending IGMP Packets....1040
13.6.3 Configuring the Multicast Service on a Cascading Network....1041
13.6.4 Configuring the Multicast Service in an MSTP Network....1043
13.6.5 Configuring the PIM Service....1045
13.6.6 Configuring the Layer 2 Reverse Multicast Service....1047
13.6.7 Configuring the Layer 3 Reverse Multicast Service....1050
13.7 Multicast Maintenance and Diagnosis....1054
13.7.1 Multicast Emulation....1054
13.7.1.1 Introduction....1054
13.7.1.2 Principles....1055
13.7.1.3 Usage Scenario....1057
13.7.1.4 Configuration....1064
13.7.1.5 Multicast Emulation Reference Standards and Protocols....1065
13.7.2 Video Quality Monitoring....1066
13.7.2.1 Introduction....1066
13.7.2.2 Basic Concepts....1067
13.7.2.3 Principles....1068

 Copyright © Huawei Technologies Co., Ltd.

13.7.2.4 Configuring Video Quality Monitoring....1071
13.7.3 RTP Video Quality Monitoring....1072
13.7.4 eMDI Video Quality Monitoring for the Multicast Service....1076
13.7.4.1 Introduction....1076
13.7.4.2 eMDI Video Quality Monitoring Indicators....1078
13.7.4.3 Principle....1081
13.7.4.4 Configuring eMDI Video Quality Monitoring....1084
13.7.5 eMDI Video Quality Monitoring for the Unicast Service....1086
13.7.5.1 Introduction to eMDI for Unicast Services....1086
13.7.5.2 eMDI Monitoring Indicators for Unicast Services....1089
13.7.5.3 eMDI Interaction Process for Unicast Services....1094
13.7.5.4 Network Application of eMDI for Unicast Services....1095
13.7.5.5 Configuring eMDI Video Quality Monitoring....1097
13.7.6 Common Multicast Maintenance Methods....1100
13.8 Multicast Standards and Protocols Compliance....1103
13.9 Multicast Feature Updates....1104
14 Clock and Time....1105
14.1 Clock and Time Feature Updates....1106
14.2 Clock and Time Overview....1106
14.2.1 Network Synchronization Requirements....1106
14.2.2 Synchronization Overview....1107
14.2.3 Clock Synchronization....1108
14.2.4 Time Synchronization....1110
14.3 Physical Layer Clock/Time Synchronization....1112
14.3.1 Physical Layer Clock/Time Synchronization Principles....1112
14.3.1.1 Synchronous Ethernet....1112
14.3.1.2 External Time Source....1113
14.3.1.3 Clock Mode....1115
14.3.1.4 Clock Source Selection....1115
14.3.2 Physical Layer Clock/Time Synchronization Usage Scenarios....1118
14.3.2.1 Applications of Clock Output....1118
14.3.2.2 Clock Synchronization of the Synchronous Ethernet Service....1118
14.3.3 Configuring the Physical Clock....1119
14.3.3.1 Configuring the System Clock Based on the Priority....1119
14.3.3.2 Configuring the System Clock Based on the SSM Clock Source Selection Mode....1120
14.3.3.3 Configuring External Clock....1123
14.3.4 Physical Layer Clock/Time Synchronization Standards and Protocols Compliance....1124
14.4 1588v2 (based on IEEE)....1125
14.4.1 Why Is 1588v2 Required....1125
14.4.2 1588v2 Basic Concepts....1126
14.4.2.1 Device Model....1126
14.4.2.2 Clock Domain and Clock ID....1128

 Copyright © Huawei Technologies Co., Ltd.

14.4.2.3 1588v2 Messages 1128
14.4.3 1588v2 Principle 1134
14.4.3.1 Clock Synchronization Mechanism 1134
14.4.3.2 BMC Algorithm 1136
14.4.3.3 Delay Compensation Mechanism 1139
14.4.4 1588v2 Network Application 1140
14.4.4.1 Recommended: Synchronization Application (Network-wide 1588v2 and Synchronous Ethernet Deployment) 1140
14.4.4.2 Synchronization Application (Network-wide 1588v2 Deployment) 1141
14.4.4.3 Synchronization Application (Clock or Time Signal Injection from an OLT) 1142
14.4.4.4 Time and Clock Synchronization Protection Application 1143
14.4.5 Configuring the 1588v2 Function 1145
14.4.6 Configuring 1588v2-related Delay Compensation for Asymmetric Fibers 1150
14.4.7 1588v2 Maintenance and Diagnosis 1152
14.4.8 1588v2 (based on IEEE) Reference Standards and Protocols 1152
14.5 1588v2 (Based on G.8275.1) 1153
14.5.1 Basic Concept 1153
14.5.2 BMCA Algorithm 1154
14.5.3 Configuring 1588v2 (Based on G.8275.1) 1155
14.5.4 1588v2 (Based on G.8275.1) Reference Standards and Protocols 1157
14.6 1588 ACR 1158
14.6.1 Why Is 1588 ACR Required 1158
14.6.2 1588 ACR Basic Concepts 1159
14.6.2.1 1588 ACR Messages 1159
14.6.2.2 PDV 1160
14.6.2.3 1588 ACR Modes 1161
14.6.2.4 1588 ACR Clock Source Selection Algorithm 1165
14.6.3 1588 ACR Principles 1166
14.6.3.1 1588 ACR Unicast Negotiation 1167
14.6.3.2 1588 ACR Principles 1168
14.6.4 1588 ACR Deployment Requirements 1169
14.6.5 1588 ACR Networking 1169
14.6.6 Configuring 1588 ACR 1171
14.6.7 1588 ACR Maintenance and Diagnosis 1175
14.6.8 1588 ACR Standard and Protocol Compliance 1175
14.6.9 Remote 1588 Packet Capture 1176
14.7 NTP 1176
14.7.1 NTP Introduction 1176
14.7.2 NTP Principle 1177
14.7.2.1 NTP Network Architecture 1177
14.7.2.2 NTP Operating Mode 1178
14.7.2.3 NTP Working Principle 1182
14.7.2.4 NTP Security Mechanism 1183

 Copyright © Huawei Technologies Co., Ltd.

14.7.3 Configuring the Network Time ... 1185
14.7.3.1 (Optional) Configuring NTP Authentication ... 1186
14.7.3.2 Configuring the NTP Broadcast Mode ... 1187
14.7.3.3 Configuring the NTP Multicast Mode ... 1190
14.7.3.4 Configuring the Unicast NTP Client ... 1193
14.7.3.5 Configuring the NTP Peer ... 1195
14.7.3.6 Configuring the NTP Multicast Mode ... 1197
14.7.3.7 (Optional) Configuring NTPv4 Control Signaling KOD ... 1200
14.7.4 NTP Standards and Protocols Compliance ... 1202

15 QoS ... 1203

15.1 Introduction to QoS ... 1204
15.2 QoS Models ... 1204
15.3 QoS Scheme ... 1206
15.4 QoS Processing ... 1208
15.5 Traffic Classification ... 1211
15.5.1 Introduction to Traffic Classification ... 1211
15.5.2 Implementation Principle of Traffic Classification ... 1211
15.5.3 Configuring the Traffic Classification ... 1214
15.6 Priority Marking ... 1215
15.6.1 Introduction to Priority Marking ... 1215
15.6.2 Basic Concepts of Priority Marking ... 1216
15.6.3 Priority Sources ... 1219
15.6.4 Implementation Principle of Priority Marking ... 1221
15.6.4.1 Priority Processing for Data Packets ... 1222
15.6.4.2 Priority Processing for DHCP & PPPoE Packets ... 1223
15.6.4.3 Priority Processing for DHCPv6 Packets ... 1225
15.6.4.4 Priority Processing for IGMP Packets ... 1226
15.6.4.5 Priority Processing for ARP Packets ... 1228
15.6.4.6 Priority Processing for ND Packets ... 1230
15.6.5 Configuring the Priority Processing ... 1230
15.7 Traffic Policing ... 1232
15.7.1 Introduction to Traffic Policing ... 1232
15.7.2 Basic Concepts of Traffic Policing ... 1232
15.7.3 Implementation Principle: CAR ... 1234
15.7.4 Traffic Policing Mode ... 1239
15.7.5 Configuring the Traffic Policing ... 1245
15.7.5.1 Configuring Rate Limitation Based on Service Port ... 1245
15.7.5.2 Configuring Rate Limitation Based on Port+CoS ... 1248
15.7.5.3 Configuring User-based Rate Limitation ... 1248
15.7.5.4 Configuring Rate Limitation Based on Port+VLAN ... 1250
15.7.5.5 Configuring GPON Rate Limitation ... 1252
15.7.5.6 Configuring Rate Limiting Based on Priority Group ... 1255

 Copyright © Huawei Technologies Co., Ltd.

15.8 Traffic Shaping....1256
15.8.1 Introduction to Traffic Shaping....1256
15.8.2 Basic Concepts of Traffic Shaping....1256
15.8.3 Traffic Shaping Implementation....1257
15.8.4 Configuring Traffic Shaping....1258
15.8.4.1 Configuring a Rate Limit for an ONT....1258
15.8.4.2 Configure Queue-based Traffic Shaping for an ONT....1259
15.8.4.3 Configuring Rate Limit for an Ethernet Port....1260
15.9 Congestion avoidance....1260
15.9.1 Introduction to Congestion avoidance....1260
15.9.2 Basic Concepts of Congestion avoidance....1261
15.9.3 Implementation Principle....1262
15.9.4 Configuring the Congestion Avoidance....1265
15.10 Congestion Management....1269
15.10.1 Introduction to Congestion Management....1269
15.10.2 Basic Concepts of Congestion Management....1269
15.10.3 Implementation Principle of Congestion Management....1270
15.10.4 Configuring the Congestion Management....1273
15.11 ACL....1275
15.11.1 Overview of ACL....1275
15.11.2 Basic Concepts of ACL....1276
15.11.3 ACL Rule Matching Sequence....1277
15.11.4 ACL Rule Matching Process....1279
15.11.5 Matching Principle for the User-defined ACL Rule....1282
15.11.6 Configuring Traffic Management Based on ACL Rules....1284
15.11.6.1 Configuration Differences Between IPv4 ACLs and IPv6 ACLs....1284
15.11.6.2 Configuring an ACL Rule....1284
15.11.6.3 Configuring ACL Matching for PPPoE Packets....1290
15.11.6.4 Controlling the Traffic Matching an ACL Rule....1293
15.11.6.5 Adding a Priority Tag to the Traffic Matching an ACL Rule....1294
15.11.6.6 Enabling the Statistics Collection of the Traffic Matching an ACL Rule....1294
15.11.6.7 Enabling the Mirroring of the Traffic Matching an ACL Rule....1295
15.11.6.8 Enabling the Redirection of the Traffic Matching an ACL Rule....1296
15.12 ACLv6....1297
15.12.1 Comparison Between ACLv6 and ACLv4....1297
15.13 HQoS....1297
15.13.1 Overview of HQoS....1297
15.13.2 Open Access of HQoS....1298
15.13.3 Basic Concepts of HQoS....1301
15.13.4 HQoS Service Model (Based on Port+VLAN)....1301
15.13.5 HQoS Service Model (Based on a CAR Group)....1302
15.13.6 HQoS Service Model (xPON Board)....1304

 Copyright © Huawei Technologies Co., Ltd.

15.13.7 HQoS Service Model (P2P Board)....1305
15.13.8 Implementation Principle of HQoS....1305
15.13.8.1 Implementation Principle of HQoS based on Port+VLAN....1305
15.13.8.2 Implementation Principle of CAR-Group-based HQoS....1308
15.13.8.3 HQoS Implementation Principle on an xPON Board....1312
15.13.8.4 HQoS Implementation Principle on a P2P Board....1315
15.13.9 Networking Application of HQoS....1318
15.13.10 Reference Standards and Protocols of HQoS....1322
15.13.11 Configuring HQoS....1322
15.13.11.1 Configuring HQoS Based on Port+VLAN....1322
15.13.11.2 Configuring HQoS Based on CAR Group....1324
15.13.11.3 Configuring HQoS for an xPON Board....1326
15.13.11.4 Configuring HQoS for a P2P Board....1330
15.14 End-to-End QoS....1334
15.14.1 FTTH End-to-End QoS Policy....1334
15.14.2 FTTB/FTTC End-to-End QoS Policy....1339
15.14.3 QoS Solution for FTTH....1341
15.14.4 QoS Solution for FTTB/FTTC....1344
16 Reliability....1348
16.1 Reliability Solution Overview....1349
16.2 xPON Type B Protection....1351
16.2.1 Introduction to xPON Type B Protection....1351
16.2.2 Basic Concepts of xPON Type B Protection....1353
16.2.3 Single-Homing xPON Type B Protection Principles....1362
16.2.4 Dual-Homing xPON Type B Protection Principles....1364
16.2.5 Configuring xPON Type B Single-Homing Protection....1368
16.2.6 Configuring xPON Type B Dual-Homing Protection....1370
16.2.7 xPON Type B Protection Reference Standards and Protocols....1373
16.3 xPON Type C Protection....1374
16.3.1 Introduction to xPON Type C Protection....1375
16.3.2 Basic Concepts of xPON Type C Protection....1376
16.3.3 Single-Homing xPON Type C Protection Principles....1385
16.3.4 Dual-Homing xPON Type C Protection Principles....1389
16.3.5 Configuring xPON Type C Single-Homing Protection....1394
16.3.6 Configuring xPON Type C Dual-Homing Protection....1396
16.3.7 xPON Type C Protection Reference Standards and Protocols....1397
16.4 Redundancy Backup of Control Boards....1398
16.4.1 Introduction to Control Board Redundancy Backup....1399
16.4.2 Principle of Active/Standby Mode....1400
16.4.3 Principle of Load-sharing Mode....1404
16.5 Ethernet Port Protection Group....1405
16.5.1 Introduction to Protection Group of Ethernet Ports....1405

 Copyright © Huawei Technologies Co., Ltd.

16.5.2 Principle of Timedelay Protection....1406
16.5.3 Configuring a Timedelay Protection Group....1408
16.6 Ethernet Link Aggregation....1409
16.6.1 Introduction to Ethernet Link Aggregation....1409
16.6.2 Principle of LACP....1410
16.6.3 Introduction to Load Bearing....1411
16.6.4 Introduction to LAG Type....1413
16.6.5 Introduction to Port Type....1415
16.6.6 Principle of LAG Setup Process....1416
16.6.7 Principle of LAG Link Switchover and Reversion....1418
16.6.8 Configuring Ethernet Link Aggregation....1419
16.6.9 Ethernet Link Aggregation Reference Standards and Protocols Compliance....1422
16.7 MSTP....1422
16.7.1 Introduction to MSTP....1422
16.7.2 Principle of MSTP....1423
16.7.3 Configuring the MSTP....1431
16.7.4 MSTP Reference Standards and Protocols....1435
16.8 ERPS....1435
16.8.1 Introduction to ERPS....1436
16.8.2 Basic Concepts of ERPS....1437
16.8.3 Principle of ERPS....1440
16.8.4 Configuring ERPS....1444
16.8.5 ERPS Reference Standards and Protocols....1448
16.9 VRRP Snooping....1449
16.9.1 Introduction to VRRP Snooping....1449
16.9.2 Principle of VRRP Snooping....1449
16.9.3 Configuring VRRP Transparent Transmission in the S+C Forwarding Mode....1451
16.9.4 VRRP Snooping Reference Standards and Protocols....1452
16.10 Ring Check....1452
16.10.1 Introduction to Ring Check....1452
16.10.2 Principle of Ring Check....1453
16.10.3 Configuring the Ring Network Detection....1456
16.11 Monitor Link....1457
16.11.1 Introduction to Monitor Link....1457
16.11.2 Principles of Monitor Link....1458
16.11.3 Configuring the Monitor Link Group....1459
17 802.1X Authentication....1461
17.1 Why Is 802.1X Authentication Required?....1461
17.2 Device Roles Defined in 802.1X....1461
17.3 Port-based Authentication....1462
17.4 802.1X Information Exchange....1464
17.5 Configuring 802.1X Authentication....1476

 Copyright © Huawei Technologies Co., Ltd.

17.5.1 ONU Functioning as the 802.1x Authentication Device....1476
17.5.2 OLT Functioning as the 802.1X Authentication Device (Configured on the CLI)....1480
17.5.3 Core Switch Functioning as the 802.1X Authentication Device....1484
18 AAA....1485
18.1 RADIUS....1487
18.2 HWTACACS....1489
18.3 Configuring the Local AAA....1492
18.4 Configuring the Remote AAA (RADIUS Protocol)....1494
18.5 Configuration Example of the RADIUS Authentication and Accounting....1501
18.6 Configuring the Remote AAA (HWTACACS Protocol)....1504
18.7 Configuration Example of the RADIUS AAA (802.1X access user)....1508
18.8 Configuration Example of HWTACACS Authentication (Management User)....1511
19 Device Management....1515
19.1 ANCP....1516
19.1.1 ANCP Introduction....1516
19.1.2 ANCP Principle....1517
19.1.3 Configuring ANCP....1521
19.1.4 ANCP Reference Standards and Protocols....1525
19.2 ANCP Proxy....1525
19.2.1 ANCP Proxy Introduction....1525
19.2.2 ANCP Proxy Principle....1527
19.2.3 Configuring ANCP Proxy....1529
19.3 LLDP....1531
19.3.1 Introduction....1531
19.3.2 LLDP Reference Standards and Protocols....1532
19.3.3 Basic Concepts....1532
19.3.4 Principles....1535
19.3.5 Network Application....1537
19.3.6 Configuring LLDP....1539
19.4 NETCONF....1541
19.4.1 Feature Updates....1541
19.4.2 Introduction....1541
19.4.3 Principles....1543
19.4.4 Networking Scenario....1545
19.4.5 Configuring NETCONF....1546
19.4.6 NETCONF Reference Standards and Protocols....1548
19.5 Telemetry....1548
19.5.1 Telemetry Feature Updates....1548
19.5.2 Introduction to Telemetry....1549
19.5.3 Telemetry Data Collection and Reporting Process....1550
19.5.4 Telemetry Data Model....1551
19.5.5 Key Telemetry Technologies....1552

 Copyright © Huawei Technologies Co., Ltd.

19.5.5.1 UDP Transport Protocol 1552
19.5.5.2 GPB Encoding Format 1556
19.5.6 Telemetry Configurations 1558
19.5.6.1 Enabling NETCONF 1558
19.5.6.2 Configuring the Data to Be Sampled 1559
19.5.6.3 Configuring a Destination Collector 1560
19.5.6.4 Creating a Subscription 1562
19.6 Inband Management VPN 1565
19.6.1 Introduction to Inband Management VPN 1565
19.6.2 Principles to Inband Management VPN 1566
19.7 Remote Software Commissioning (Ethernet) 1567
19.7.1 Introduction 1567
19.7.2 Principles (Based on DHCP) 1570
19.7.3 Configuring NAC-based Remote Software Commissioning for Ethernet Upstream Transmission 1572
19.7.4 Remote Software Commissioning (Ethernet) Reference Standards and Protocols 1578
19.8 OLT ZTD 1578
19.8.1 Feature Updates 1579
19.8.2 Why the OLT Needs to Support ZTD 1579
19.8.3 OLT ZTD Application Scenarios 1580
19.9 Power Saving 1582
19.9.1 Introduction to Energy Saving 1583
19.9.2 Energy Saving Principles 1583
19.9.2.1 Energy Saving Principles 1583
19.9.2.2 Energy Saving Technology Description 1586
19.9.2.2.1 Automatic Shutdown of Unused High-speed Interfaces 1586
19.9.2.2.2 Automatic Zone-based Fan Speed Adjustment 1586
19.9.2.2.3 Automatic Shutdown of Unused Ports 1588
19.9.2.2.4 Automatic Shutdown of Unused Boards 1589
19.9.2.2.5 Automatic Shutdown of Unused Channels 1590
19.9.2.2.6 Dynamic Frequency Adjustment 1591
19.9.2.2.7 Intelligent Board Scheduling 1591
19.9.3 Energy Saving Modes and Constraints 1592
19.9.4 Principles of Power Savings and Power Estimation 1592
19.9.5 Energy Saving Configuration Description 1593
19.9.5.1 Energy Saving Configuration 1593
19.9.5.1.1 Configuring an Energy Saving Mode 1593
19.9.5.1.2 Configuring Custom Energy Saving Measures 1595
19.9.5.1.3 Querying Energy Saving Configurations 1596
19.9.5.1.4 Powering Off an Unused Port/Board 1599
19.9.5.2 Querying Energy Saving Results 1599
19.9.5.2.1 Querying Energy Saving Status 1599
19.9.5.2.2 Querying the Shutdown Status of a Board 1600

 Copyright © Huawei Technologies Co., Ltd.

19.9.5.2.3 Querying the Shutdown Status of a Port....1600
19.9.5.2.4 Querying the Shutdown Status of a Channel....1601
19.9.5.2.5 Querying Zone-based Fan Speed Adjustment Results....1601
19.9.5.2.6 Querying Dynamic Frequency Adjustment Results....1601
19.9.5.2.7 Querying Intelligent Board Scheduling Results....1602
19.9.5.3 Energy Saving Configuration Examples....1603
19.9.5.3.1 Configuring the Deep Energy Saving Mode on an OLT....1603
19.9.5.3.2 Configuring the Deep Energy Saving Mode on the NMS....1604
19.9.6 Energy Saving Conditions....1607
19.9.6.1 Function Conditions....1607
19.9.6.1.1 Energy Saving Mode Conditions....1607
19.9.6.1.2 Energy Saving Measure Conditions....1608
19.9.6.1.3 Energy Saving Parameter Conditions....1610
19.9.6.2 Board Conditions....1610
19.9.6.2.1 Energy Saving Configuration Conditions....1610
19.9.6.2.2 Support for Energy Consumption Visualization....1611
19.9.7 FAQ....1612
19.9.7.1 Can the Custom Energy Saving Mode Be Configured on a Device?....1612
19.9.7.2 Does a Custom Energy Saving Measure Take Effect Immediately After Configuration?....1612
19.9.7.3 Link for the Manual of Configuring Custom Energy Saving Measures on the NMS....1613
19.10 Version and Data Management....1613
19.10.1 Introduction....1613
19.10.2 Principle....1613
19.10.3 ISSU....1615
19.11 Anti-theft....1617
19.11.1 Anti-theft Feature Updates....1617
19.11.2 Anti-theft Introduction....1617
19.11.3 Anti-theft Principles....1618
19.11.4 Configuring Anti-theft....1620
19.12 Environment Monitoring....1620
19.12.1 EMUB Monitoring Solution....1620
19.12.1.1 DCDU-12E PDB....1620
19.12.1.2 EMUB Monitoring Unit....1623
19.12.1.3 Checking the Environment Monitoring Cables of the EMUB....1633
19.12.1.4 Commissioning the EMUB....1635
19.12.2 Power3000 Monitoring Solution....1638
19.12.2.1 Monitoring Module CPMU01 of the Power3000....1638
19.12.2.2 Checking the Environment Monitoring Cable of the Power3000....1639
19.12.2.3 Configuring the Power3000 Environment Monitoring Mode....1640
19.12.3 ETP4890 Monitoring Solution (AC Power Supply)....1643
19.12.3.1 ETP4890 Power System....1643
19.12.3.2 PMIB02 Sensor Transfer Box....1650

 Copyright © Huawei Technologies Co., Ltd.

19.12.3.3 Checking Environment Monitoring Cables....1653
19.12.3.4 Configuring the Environment Monitoring Parameters of the ETP4890....1655
19.12.4 ETP48200 Monitoring Solution....1659
19.12.4.1 ETP48200-B2A1 Power System....1659
19.12.4.2 Checking the Environment Monitoring Cable....1669
19.12.4.3 Configuring the Environment Monitoring Parameters of the ETP48200....1671
19.12.5 Fan Tray Monitoring Solution....1675
19.12.5.1 Fan Tray Monitoring Principles....1675
19.12.5.2 MA5800-X17 Fan Tray....1676
19.12.5.3 MA5800-X15 Fan Tray....1679
19.12.5.4 MA5800-X7 Fan Tray....1682
19.12.5.5 MA5800-X2 Fan Tray....1685
19.12.5.6 Commissioning the Fan EMU....1687
19.13 RTU....1689

20 Slicing....1691

20.1 Slicing Solution Change Description....1691
20.2 Slicing Solution Overview....1692
20.2.1 Background of Slicing....1692
20.2.2 Benefits of the Slicing Solution....1694
20.2.3 Architecture of the Slicing Solution....1695
20.2.4 Implementation Principles of Access Network Slicing....1697
20.2.5 Key Technologies....1702
20.2.5.1 HDBA....1702
20.2.6 Typical Application Scenarios of the Slicing Solution....1704
20.2.6.1 Application Networking of the Slicing Solution....1704
20.2.6.2 Customized Network Services for Dedicated Network Slices....1705
20.2.6.3 Differentiated Service Experience of Service Slices....1707
20.3 Slice Lifecycle Management....1709
20.4 Slice Deployment and Configuration....1710
20.4.1 Dedicated Network Slice Deployment....1710
20.4.1.1 Example of Hard-Isolated Dedicated Network Slice Deployment (NCE)....1710
20.4.1.2 Example of Hard-Isolated Dedicated Network Slice Deployment (CLI)....1714
20.4.2 Service Slice Deployment....1721
20.4.2.1 Service Slice Deployment Example (NCE)....1721
20.4.2.2 Service Slice Deployment Example (CLI)....1728
20.5 Reference....1730
20.5.1 Parameter Reference for a Service Slice SLA Template....1731
20.5.2 Acronyms, Abbreviations, and Basic Concepts....1733
20.5.3 Reference Standards and Protocols for Slicing....1734

21 Aggregation Management....1735

21.1 Changes in Aggregation Management....1735
21.2 Hardware Version Description for Aggregation Management....1735

 Copyright © Huawei Technologies Co., Ltd.

21.3 Introduction to the Aggregation Management Solution....1736
21.4 Basic Concepts of Aggregation Management....1738
21.5 Implementation Principle of Aggregation Management....1740
21.6 Application Scenario of Aggregation Management....1743
21.7 Aggregation Management Configuration....1744
21.7.1 Digital Certificate Loading Process....1744
21.7.2 Typical Networking and Configuration Process....1745
21.7.3 Configuration Example....1749
21.7.3.1 Customized Configuration Example....1749
21.7.3.2 Batch Configuration Example....1752

 Copyright © Huawei Technologies Co., Ltd.

1 Feature Specifications and Limitations

For carrier user and in online mode, enable the **Specifications Query Tool** to see detailed feature specifications and limitations.

For enterprise user and in online mode, enable the **Specifications Query Tool** to see detailed feature specifications and limitations.

This tool has 3 functions:

- Specification query: supports specification query for different products and different versions.
- Specification comparison: supports feature specification comparison between different boards of the same product without different versions, or between different boards of the same product with the same version.
- Specification list export: the specification list of a single product or a single board of a version can be exported.

 Copyright © Huawei Technologies Co., Ltd.

2 GPON

Gigabit passive optical network (GPON) is a PON technology that is standardized by the ITU-T Recommendations G.984.x. A GPON device supports high-bandwidth transmission. GPON effectively solves the bandwidth bottleneck problem in the twisted-pair access and meets users demands on high-bandwidth services.

2.1 Introduction to GPON

2.2 GPON Working Principle

2.3 GPON System Principle

2.4 Key GPON Techniques

2.5 GPON Terminal Authentication and Management

2.6 Continuous-Mode ONU Detection
This function is used to detect the continuous-mode ONU of terminal devices such as the MDU and ONT.

2.7 Remote Software Commissioning (PON)

2.8 GPON Configuration Guide (Profile mode)
GPON configurations include the configurations on GPON profiles, ONTs, and ports. The following section describes configuration methods.

2.9 GPON Configuration Guide (Distributed mode)
GPON configurations include the configurations on GPON profiles, ONTs, and ports. The following section describes configuration methods.

2.10 GPON Reference Standards and Protocols

2.1 Introduction to GPON

What Is GPON

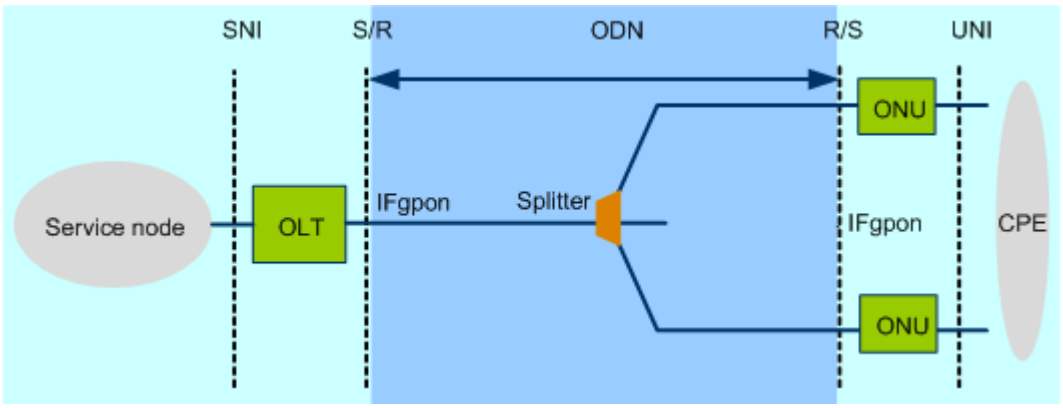
Passive Optical Network (PON) is a point to multi-point (P2MP) passive optical network.

Gigabit passive optical network (GPON) is a PON technology that is standardized by the ITU-T Recommendations G.984.x.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-1 shows a GPON network architecture.

Figure 2-1 GPON network architecture



IFgpon: GPON interface	SNI: Service Node Interface
UNI: User to Network Interface	CPE: Customer Premises Equipment
S/R: Optical line termination interface to optical trunk line	R/S: Optical network unit interface to optical distribution network

- The optical line terminal (OLT) is an aggregation device located at the central office (CO) for terminating the PON protocol.
- Optical network units (ONUs)/Optical network terminal (ONTs) are located on the user side, providing various ports for connecting to user terminals. The OLT and ONUs communicate with each other through the optical distribution network (ODN).
- The optical distribution network (ODN) is composed of passive optical components (POS), such as optical fibers, and one or more passive optical splitters. The ODN provides optical channels between the OLT and ONUs. It interconnects the OLT and ONUs and is highly reliable.

  NOTE

  The ODN network is passive, indicating that no optical amplifier or regenerator is deployed on the ODN network, thereby reducing maintenance costs of outdoor devices.

Why Is GPON Required

As the wide use of broadband services and fiber-in and copper-out development, carriers require a longer transmission reach, higher bandwidth, reliability, and lower operating expense (OPEX) on services. GPON supports the following functions to meet these requirements:

- Longer transmission distance: The transmission media of optical fibers covers up to 60 km coverage radius on the access layer, resolving transmission distance and bandwidth issues in twisted pair transmission.

 Copyright © Huawei Technologies Co., Ltd.

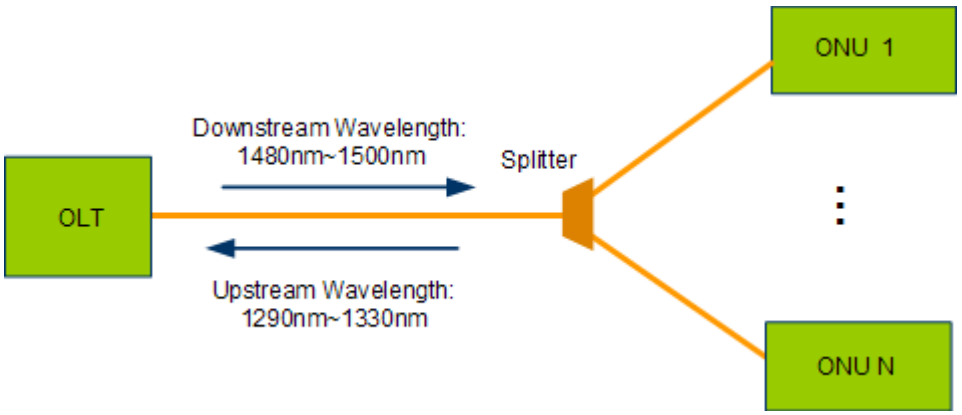
- Higher bandwidth: Each GPON port can support a maximum transmission rate of 2.5 Gbit/s in the downstream direction and 1.25 Gbit/s in the upstream direction, meeting the usage requirements of high-bandwidth services, such as high definition television (HDTV) and outside broadcast (OB).
- Flexible user experience on full services: Flexible QoS measures support traffic control based on users and user services, implementing differentiated service provisioning for different users.
- Optical splitting: A single optical fiber at the CO is split into multiple drop fibers, saving feeder optical fiber resources and reducing O&M costs.

2.2 GPON Working Principle

System Transmission Principle

Figure 2-2 shows the system transmission principle of the GPON network.

Figure 2-2 GPON System Transmission Principle



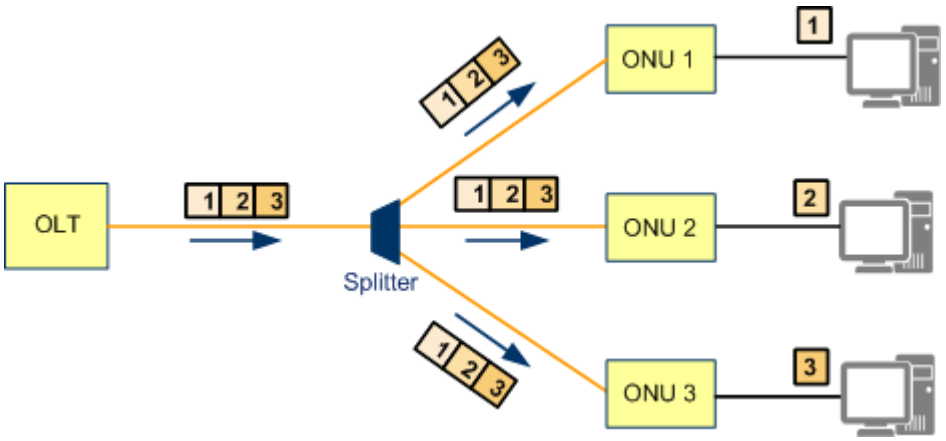
- In the GPON network, the OLT is connected to the optical splitter through a single optical fiber, and the optical splitter is then connected to ONUs. Different wavelengths are adopted in the upstream and downstream directions for transmitting data. Specifically, wavelengths range from 1290 nm to 1330 nm in the upstream direction and from 1480 nm to 1500 nm in the downstream direction.
- The GPON adopts WDM to transmit data of different upstream/downstream wavelengths over the same ODN. Data is broadcast in the downstream direction and transmitted in the TDMA mode (based on timeslots) in the upstream direction.

GPON Downstream Transmission Principle

All data is broadcast to all ONUs from the OLT. The ONUs then select and receive their respective data and discard the other data. Figure 2-3 shows the details.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-3 Downstream communication principle of GPON



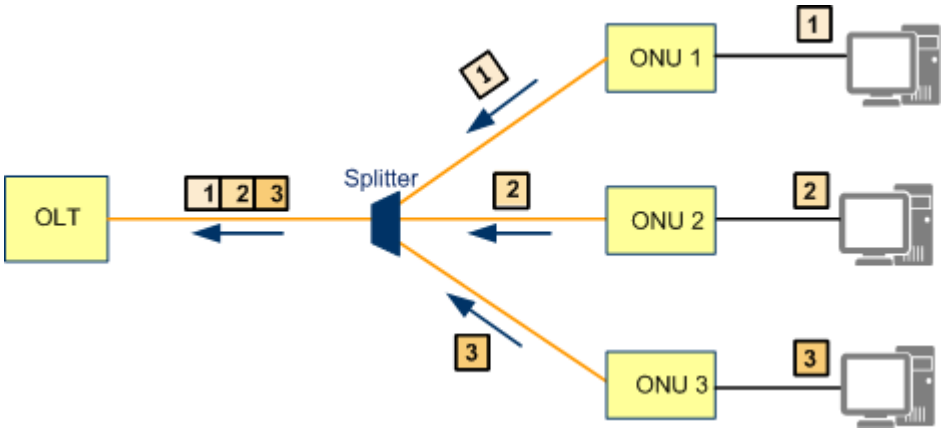
Main features:

- Supports point-to-multipoint (P2MP) multicast transmission.
- Broadcasts the same data to all ONUs and differentiates ONU data by GEM port ID.
- Allows an ONU to receive data packets intended for it and discards those not intended for it.

GPON Upstream Transmission Principle

In the upstream direction, each ONU can send data to the OLT only in the timeslot permitted and allocated by the OLT. This ensures that each ONU sends data in a given sequence, avoiding upstream data conflicts. **Figure 2-4** shows the details.

Figure 2-4 Upstream communication principle of GPON



Main features:

- Supports time division multiple access (TDMA).
- Transits data on an exclusive timeslot.

Copyright © Huawei Technologies Co., Ltd.

- Couples optical signals on an optical splitter.
- Detects and prevents collisions through ranging.

2.3 GPON System Principle

2.3.1 GPON Basic Concepts

GEM Frame

In the gigabit-capable passive optical network (GPON) system, a GPON encapsulation mode (GEM) frame is the smallest service-carrying unit and the basic encapsulation structure. All services need to be encapsulated in GEM frames and transmitted on GPON lines, which are identified by GEM ports.

- Each GEM port is identified by a unique port ID. The port ID is globally allocated by the OLT. Therefore, the ONUs connected to the same OLT GPON port cannot use GEM ports that have the same port ID.
- A GEM port is used to identify the virtual service channel that carries the service stream between the OLT and the ONU. It is similar to the virtual path identifier (VPI)/virtual channel identifier (VCI) of the asynchronous transfer mode (ATM) virtual connection.

Figure 2-5 shows the GEM frame structure.

Figure 2-5 GEM frame structure

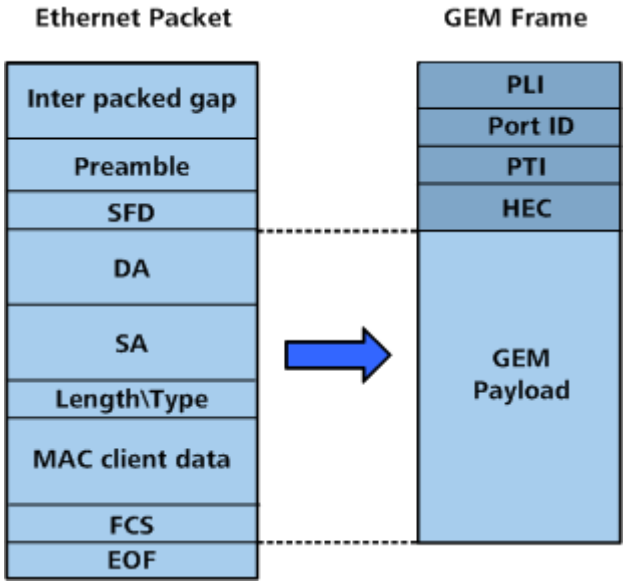
A GEM header consists of PLI, Port ID, PTI, and header error check (HEC) and is used for differentiating data of different GEM ports. The fields are described as follows:

- PLI: indicates the length of data payload.
- Port ID: uniquely identifies a GEM port.
- PTI: indicates the payload type. It is used for identifying the status and type of data that is being transmitted, for example, whether the operation, administration and maintenance (OAM) message is being transmitted and whether data transmission is complete.
- HEC: ensures the forward error correction (FEC) function and transmission quality.
- Fragment payload: indicates the data frame fragment.

The following section describes the GEM frame structure based on the mapping of the Ethernet service in GPON mode, as shown in Figure 2-6.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-6 GEM frame structure



- The GPON system parses Ethernet frames and maps data into GEM payloads for transmission.
- Header information is automatically encapsulated into GEM frames.
- The mapping format is clear and has good compatibility.

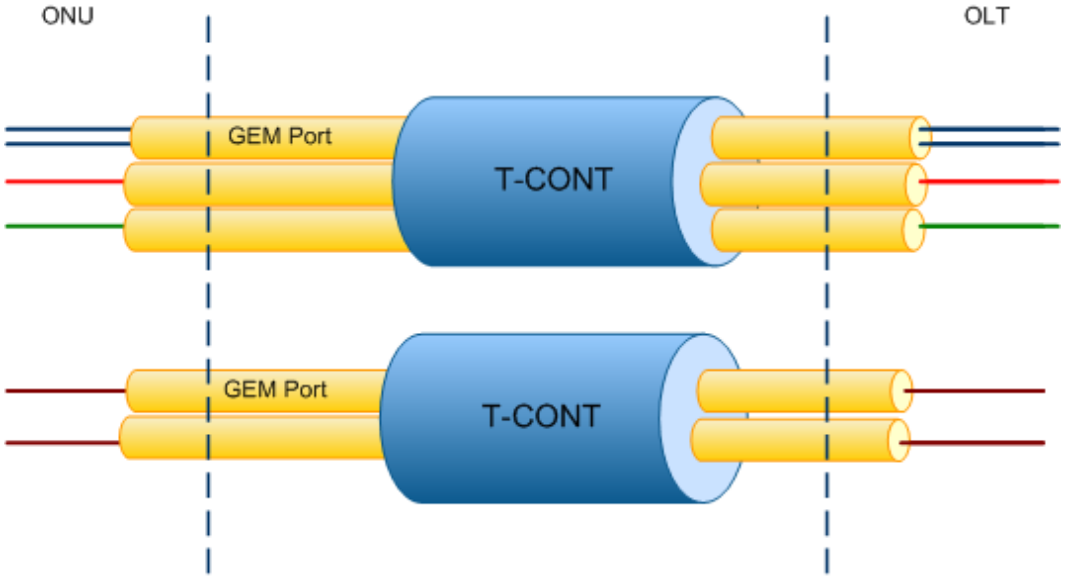
T-CONT

Transmission container (T-CONT) is a service carrier in the upstream direction in the GPON system. All GEM ports are mapped to T-CONTs. Then service streams are transmitted upstream by means of OLT's dynamic bandwidth assignment (DBA) scheduling. T-CONT is the basis of DBA implementation. An ONU applies for bandwidth using T-CONTs, and the OLT grants bandwidth through T-CONTs, thereby realizing the upstream DBA in the entire GPON system.

T-CONT is the basic control unit of the upstream service stream in the GPON system. Each T-CONT is identified by Alloc-ID. The Alloc-ID is allocated by the GPON port of the OLT, and the T-CONTs used by ONUs connected to the same GPON port of OLT cannot have the same Alloc-IDs.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-7 T-CONT structure



There are five types of T-CONT. T-CONT selection varies during the scheduling of different types of upstream service streams. Each T-CONT bandwidth type has its own quality of service (QoS) feature. QoS is mainly represented by the bandwidth guarantee, which can be classified into fixed, assured, non-assured, best-effort, and hybrid modes (corresponding to type 1 to type 5 listed in **Table 2-1**).

NOTE

In **Table 2-1**, X indicates the fixed bandwidth value, Y indicates the assured bandwidth value, Z indicates the maximum bandwidth value, and No indicates not involved.

Table 2-1 T-CONT types

Bandwidth Type	T-CONT Type				
	Type1	Type2	Type3	Type4	Type5
Fixed Bandwidth	X	-	-	-	X
Assured Bandwidth	-	Y	Y	-	Y
Maximum Bandwidth	Z=X	Z=Y	Z>Y	Z	Z≥ X + Y

Copyright © Huawei Technologies Co., Ltd.

Bandwidth Type	T-CONT Type				
	Type1	Type2	Type3	Type4	Type5
Description	• The fixed bandwidth is reserved for specific ONUs or specific services on ONUs. It cannot be used by other ONUs even if no upstream service streams are carried on the specific ONUs. • It applies to services that are sensitive to service quality. The services can be TDM or VoIP services.	• The assured bandwidth is available at any time required by an ONU. When the bandwidth required by the service streams on the ONU is smaller than the assured bandwidth, the system can use the DBA mechanism to allocate the remaining bandwidth to services on other ONUs. • Because DBA is required, this type	• **Type3** is the combination of the assured bandwidth and maximum bandwidth. The system assures some bandwidth for subscribers and allows subscribers to preempt bandwidth. However, the total used bandwidth cannot exceed the maximum configured bandwidth. • It applies to VoIP services.	• This type is the maximum bandwidth that can be used by an ONU, fully providing the bandwidth required by the ONU. • It applies to IPTV and other high-speed Internet services.	**Type5** is the combination of the fixed, assured, and maximum bandwidth. It supports the following functions: Reserves bandwidth for subscribers and the bandwidth cannot be preempted by other subscribers. Provides the bandwidth to an ONU at any time when required. Allows subscribers to preempt some bandwidth. (The total used bandwidth cannot exceed the maximum configured bandwidth.)

 Copyright © Huawei Technologies Co., Ltd.

Bandwidth Type	T-CONT Type				
	Type1	Type2	Type3	Type4	Type5
		provides a less real-time performance compared with the fixed bandwidth.			

2.3.2 GPON Service Multiplexing

GPON encapsulation mode (GEM) ports and transmission containers (T-CONTs) divide a PON network into virtual connections for service multiplexing.

- Each GEM port can carry one or more types of service stream. After carrying service streams, a GEM port must be mapped to a T-CONT before upstream service scheduling. Each ONU supports multiple T-CONTs that can have different service types.
- A T-CONT can be bound to one or more GEM ports, depending on customers' data plan. On the OLT, GEM ports are demodulated from the T-CONT and then service streams are demodulated from the GEM port payload for further processing.

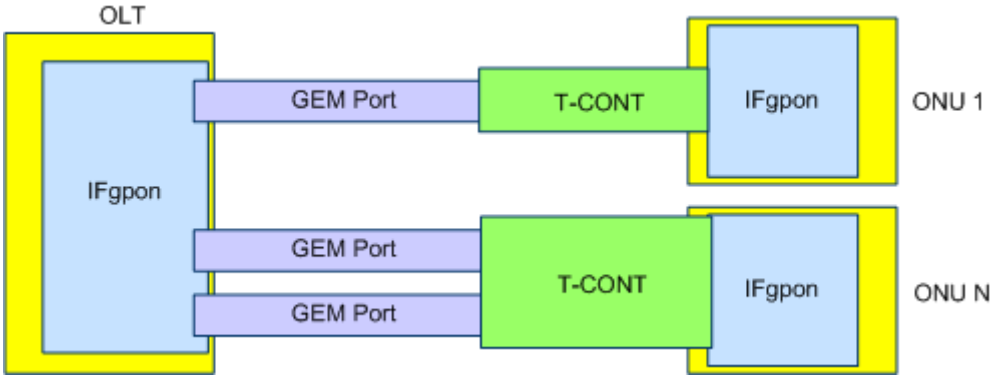
Service Mapping Relationships

- In the upstream direction,
  - An ONU sends Ethernet frames to GEM ports based on configured mapping rules between service ports and GEM ports. Then, the GEM ports encapsulate the Ethernet frames into GEM packet data units (PDUs) and add these GEM PDUs to T-CONT queues based on mapping rules between GEM ports and T-CONT queues. Then, the T-CONT queues use timeslots for upstream transmission to send GEM PDUs to the OLT.
  - The OLT receives the GEM PDUs and obtains Ethernet frames from them. Then, the OLT sends Ethernet frames from a specified uplink port based on mapping rules between service ports and uplink ports.

  Figure 2-8 shows GPON service mapping relationships in the upstream direction.

 Copyright © Huawei Technologies Co., Ltd.

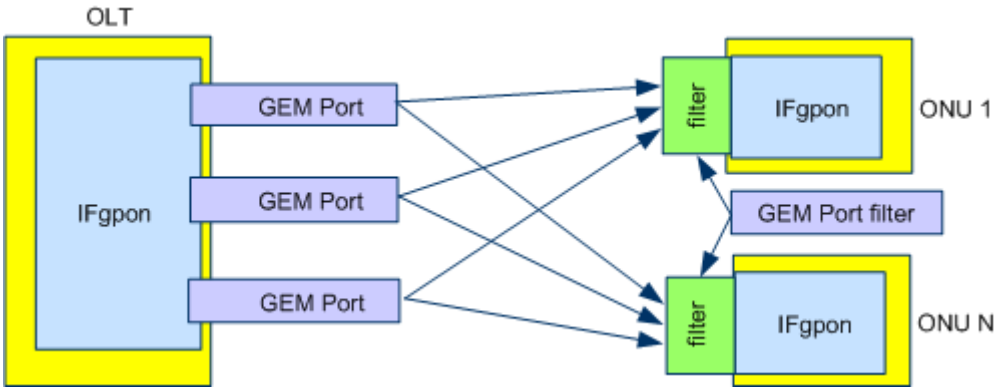
Figure 2-8 GPON service mapping relationships in the upstream direction



- In the downstream direction,
  - The OLT sends Ethernet frames to the GPON service processing module based on configured mapping rules between service ports and uplink ports. The GPON service processing module then encapsulates the Ethernet frames into GEM PDUs for downstream transmission using a GPON port.
  - GPON transmission convergence (GTC) frames containing GEM PDUs are broadcast to all ONUs connected to the GPON port.
  - The ONU filters the received data according to the GEM port ID contained in the GEM PDU header and retains the data only belonging to the GEM ports of this ONU. Then, the ONU decapsulates the data to Ethernet frames and sends them to end users using service ports.

  Figure 2-9 shows GPON service mapping relationships in the downstream direction.

Figure 2-9 GPON service mapping relationships in the downstream direction

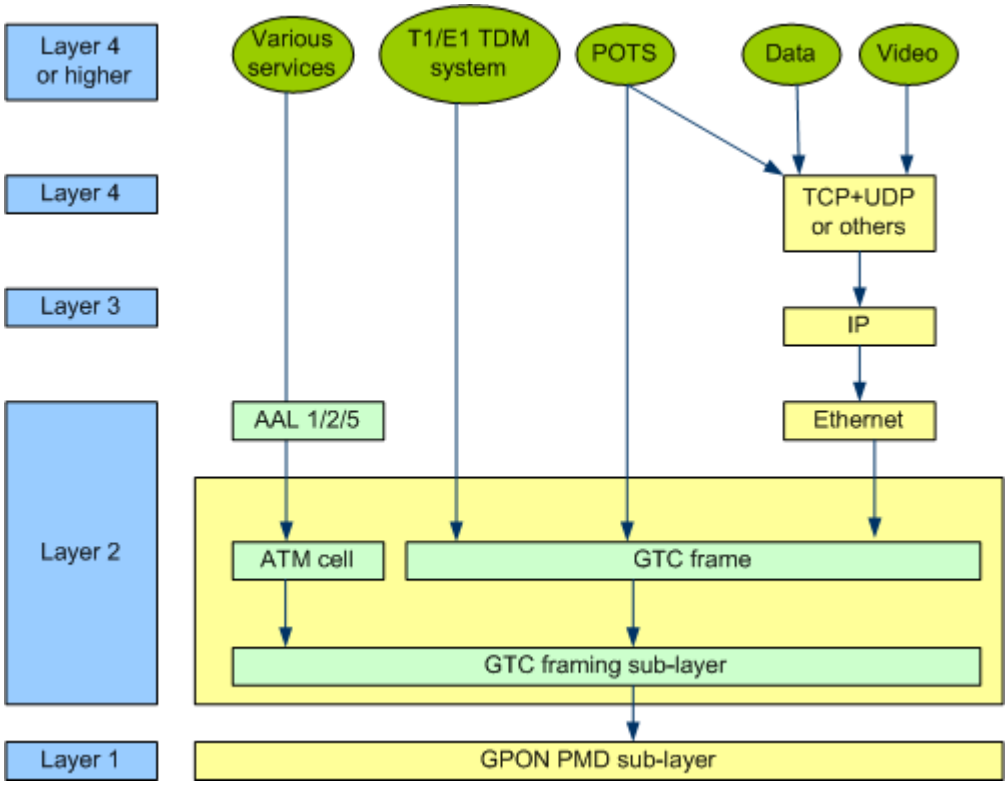


2.3.3 GPON Protocol Stacks

ITU-T Recommendation G.984.3 defines a new set of frame structures, which consider traditional voice, video, and Ethernet packets as payloads of GPON frames. **Figure 2-10** shows the structure of GPON protocol stacks.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-10 Structure of GPON protocol stacks



GPON protocol stacks involve the physical medium dependent (PMD) layer and GPON transmission convergence (GTC) layer.

PMD Layer

The GPON PMD layer corresponds to the GPON interfaces between OLTs and ONUs. Parameter values of the GPON interfaces specify the maximum reach and split ratio for a GPON system.

GTC Layer

The GTC layer is used to encapsulate payloads using ATM cells or GEM frames, and GEM frames are commonly used in GPON systems. GEM frames can carry Ethernet, POTS, E1, and T1 cells.

GTC is the core GPON layer, where media access is controlled for upstream service flows and ONUs are registered. Ethernet frame payloads are encapsulated into GEM frames and then packetized as GTC frames. These GTC frames are converted to binary codes for transmission based on interface parameters configured at the physical layer. The process is reversed on the receive end. Specifically, the receive end decapsulates the data to obtain GTC frames, GEM frames, and then payloads for data transmission.

The GTC layer is classified as TC adaptation sub-layer and GTC framing sub-layer by structure.

- The TC adaptation sub-layer involves the ATM, GEM TC, and optical network terminal management and control interface (OMCI) adapters and dynamic bandwidth assignment (DBA) control module. ATM and GEM TC adapters

 Copyright © Huawei Technologies Co., Ltd.

identify OMCI channels by virtual path identifier (VPI)/virtual channel identifier (VCI) or GEM port ID. OMCI adapters can exchange OMCI channel data with the ATM and GEM TC adapters and send the OMCI channel data to OMCI entities. The DBA control module is a common functional module, which generates ONU reports and controls DBA allocation.

- On the GTC framing sub-layer, GTC frames include GEM blocks, PLOAM blocks, and embedded OAM blocks. The GPON framing sub-layer provides the following functions:
  - Multiplexing and demultiplexing: The PLOAM part and the GEM part are multiplexed into the downlink TC frame according to the boundary information indicated by the frame header, and the PLOAM part and the GEM part can be extracted from the uplink TC frame according to the frame header indication.
  - Frame header generation and decoding: The TC frame header of the downstream frame is generated according to the format requirements, and the frame header of the upstream frame is decoded. At the same time, the embedded OAM information directly encapsulated in the GTC frame header is terminated, and is used to directly control the sub-layer.
  - Alloc-ID-based internal routing: The data from or to the GEM TC adapter is routed according to the internal representation of the Alloc-ID.

The GTC layer consists of plane C/M and plane U based on functions.

- The protocol stacks of plane C/M include embedded OAM, PLOAM, and OMCI. Embedded OAM and PLOAM channels are used for managing PMD and GTC sub-layer functions. OMCI provides a unified system for upper-layer sub-layer management.
  - Embedded OAM channels are defined in GTC frame headers for determining bandwidths, exchanging data, and dynamically allocating bandwidths.
  - Dedicated space is reserved in GTC frames for format-based PLOAM channels. The PLOAM channels carry the PMD and GTC management information that does not pass through the embedded OAM block.
  - OMCI channels are used for managing services.
- Service flows on plane U are identified based on service flow types (ATM or GEM) and port ID/VPI. Port IDs identify GEM service flows and VPIs identify ATM service flows. In T-CONTs, bandwidths are allocated and QoS is controlled using the timeslots that can be adjusted.

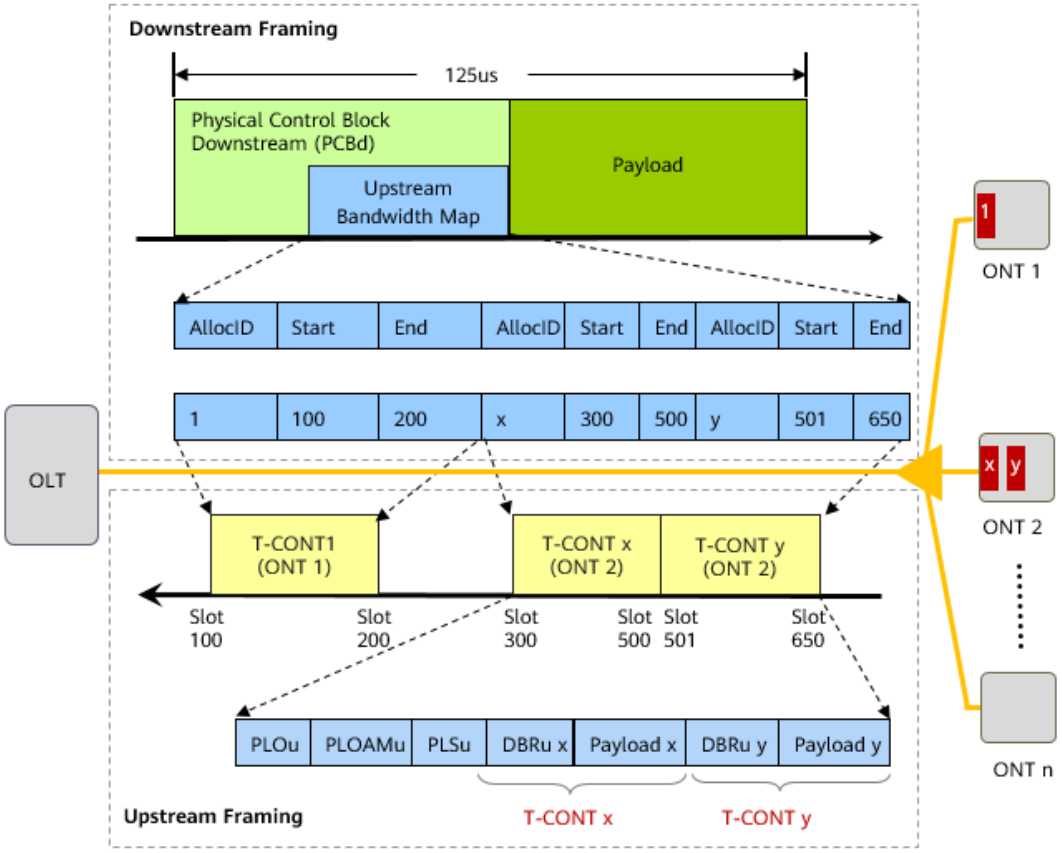
2.3.4 GPON Frame Structure

GPON Frame Structure

Figure 2-11 shows the GPON frame structure.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-11 GPON frame structure



Upstream GPON Frame

An upstream GPON frame has a fixed length of 125 μs. Each upstream frame contains the content carried by one or more T-CONTs. All ONUs connected to a GPON port share the upstream bandwidth.

- All ONUs connected to a GPON port send their data upstream at their own timeslots according to bandwidth map (BWmap) requirements.
- Each ONU reports the status of data to be sent to the OLT using upstream frames. Then, the OLT uses DBA to allocate upstream timeslots to ONUs and sends updates in each frame.

In **Figure 2-11**, an upstream GPON frame consists of the physical layer overhead upstream (PLOu), PLOAM upstream (PLOAMu), power level sequence upstream (PLSu), dynamic bandwidth report upstream (DBRu), and payload fields, as described in **Table 2-2**.

Copyright © Huawei Technologies Co., Ltd.

Table 2-2 Field description for an upstream GPON frame

Field	Description	Function
PLOu	Upstream physical layer overhead	Used for frame alignment, synchronization, and identification for an ONU.
PLOAMu	PLOAM messages of upstream data	Used for reporting ONU management messages, including maintenance and management status. This field may not be contained in a frame but must be negotiated.
PLSu	Upstream power level sequence	Used by ONUs for adjusting optical port power. This field may not be contained in a frame but must be negotiated.
DBRu	Upstream dynamic bandwidth report	Used for reporting the T-CONT status to apply for bandwidth next time and for allocating dynamic bandwidths. This field may not be contained in a frame but must be negotiated.
Payload	Payload user data	Can be a DBA status report or data frame. If this field is a data frame, this field consists of a GEM header and frames.

Downstream GPON Frame

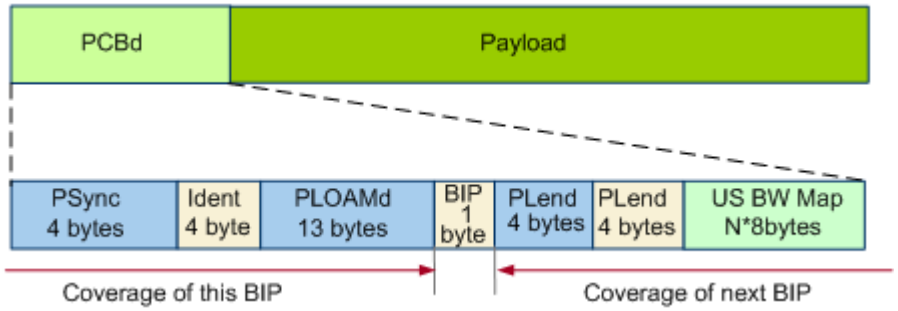
A downstream GPON frame has a fixed length of 125 μs and comprises physical control block downstream (PCBd) and payload. The OLT broadcasts PCBd to all ONUs. Then, the ONUs receive the PCBd and perform operations based on the information contained in PCBd. PCBd mainly consists of the GTC header and BWmap.

- The GTC header is used for frame delimitation, synchronization, and forward error correction (FEC).
- The BWMap field notifies every ONU of upstream bandwidth allocation. It specifies the start and end upstream timeslots for the T-CONTs of each ONU, ensuring that all ONUs send data using the timeslots specified by the OLT to prevent data conflict.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-12 shows the structure of the PCBd shown in **Figure 2-11**.

Figure 2-12 PCBd structure



PCBd contains PSync, Ident, PLOAMd, BIP, PLend, and US BW Map fields, where US BW Map is the upstream bandwidth mapping sent by the OLT for each T-CONT. **Table 2-3** describes each field.

Table 2-3 PCBd field description

Field	Description	Function
PSync	Physical synchronization domain, frame synchronization information	Used by ONUs to specify the start of each frame.
Ident	Identification domain	Used for sorting a frame in the frames of the same type in length sequence.
Downstream PLOAM (PLOAMd)	PLOAM messages of downstream data	It is used to transmit the OAM message of the physical layer to the ONU. This field may not be contained in a frame but must be negotiated.
BIP	Bit interleaved parity	Used for performing a parity check for all bytes between two BIP fields (excluding the preamble and delimit) to monitor error codes.
PLend	Length of downstream payloads	Used for specifying the length of the BWmap field.

 Copyright © Huawei Technologies Co., Ltd.

Field	Description	Function
Upstream bandwidth map (US BW Map)	Upstream bandwidth mapping	Used by the OLT for sending the upstream bandwidth mapping to each T-CONT. The BWmap specifies the start and end times for each T-CONT in transmitting data. • The StartTime field contains the 16-bit number that indicates the starting time of the allocation. **NOTE** This time is measured in bytes, starting with zero at the beginning of the upstream GTC frame. This limits the size of the upstream GTC frame to 65536 bytes. • The StopTime field contains the 16-bit number that indicates the stopping time of the allocation. **NOTE** This time is measured in bytes, starting with zero at the beginning of the upstream GTC frame. The StopTime points to the last valid data byte associated with this allocation.

2.3.5 OMCI

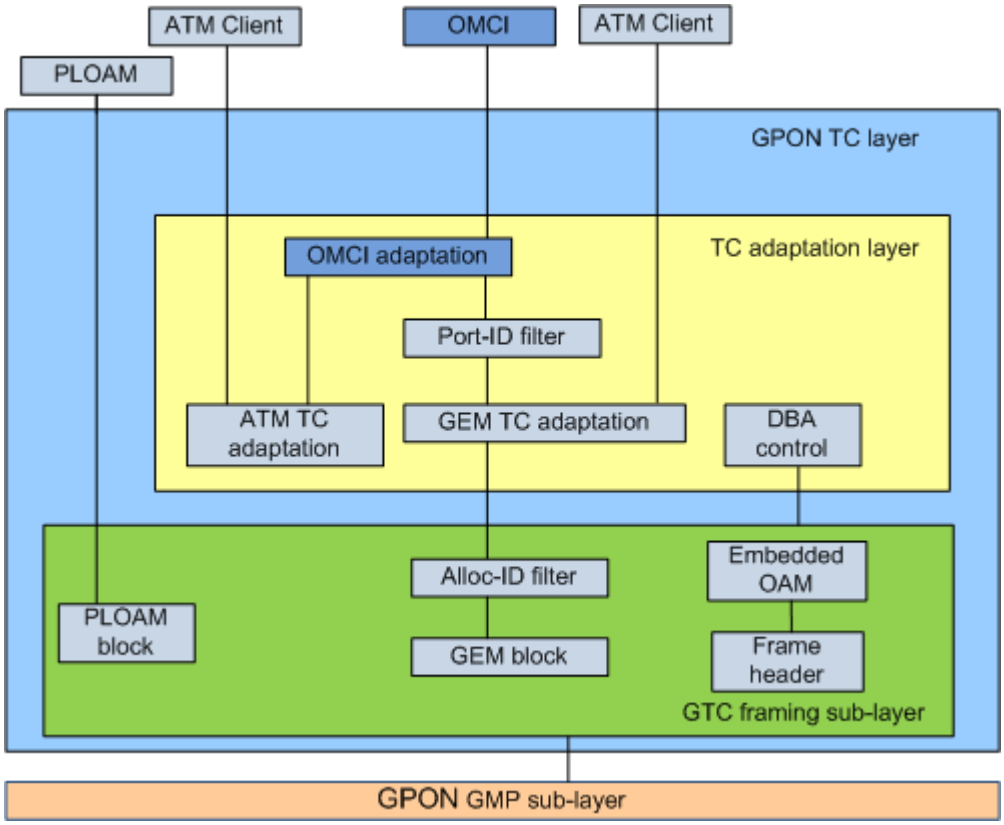
Basic Concepts

OMCI is a type of ITU-T Recommendation G.984.4-compliant configuration and transmission channel, which is used to transmit OMCI messages over dedicated ATM PVCs or GEM ports established between an OLT and an ONT. The OMCI messages are used for discovering ONTs for management and control.

 Copyright © Huawei Technologies Co., Ltd.

OMCI Position in GPON Protocol Stacks

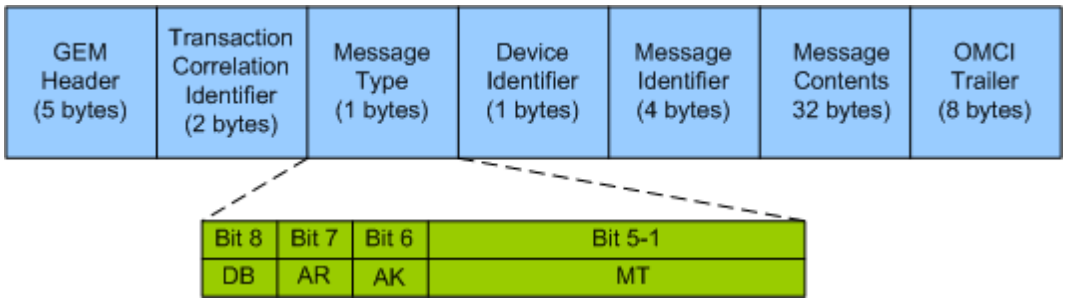
Figure 2-13 GPON Protocol Stacks



OMCI Message Format

OMCI messages are strictly limited in length and format. Specifically, the length is consistently 53 bytes and the length of the OMCI data unit is 48 bytes. For details, see **Figure 2-14**.

Figure 2-14 OMCI message format



- GEM Header: includes GEM payload, GEM port ID, payload type indicator (PTI), and header error control (HEC).

 Copyright © Huawei Technologies Co., Ltd.

- Transaction Correlation Identifier: The value of this field must be the same in a request and the response to this request. The highest order of this field indicates the priority of an OMCI message. Value 0 indicates a low priority and value 1 indicates a high priority.
- Message type:
  - DB: a destination bit, which is consistently 0.
  - AR: an acknowledge request, indicating whether an OMCI message requires the response from the peer end. Value 0 indicates that the response is not required and value 1 indicates that the response is required.
  - AK: acknowledgement, indicating whether an OMCI message is a response. Value 0 indicates not and value 1 indicates yes.
  - MT: message type, which supports up to 32 message types, including Create, Delete, Set, Get, and MIB upload. In the G.984.4 protocol, codes 4 to 28 are used, and other codes are reserved.
- Device identifier: The value of this field is consistently **0xA**.
- Message Identifier: a 2-byte entity or instance ID.
- Message Contents: packet payload.
- OMCI trailer: Two bytes are consistently 0, two bytes are packet length 0x28, and four bytes are CRCs.

OMCI Management

The OLT controls the ONU using the OMCI. The OMCI protocol allows the OLT to:

- Establish and release connections with the ONU.
- Manage the UNIs on the ONU.
- Request configuration information and performance statistics from the OSS.
- Autonomously inform the system administrator of events, such as link failures.

The OMCI protocol runs over a GEM connection between the OLT controller and the ONU controller. The GEM connection is established during ONU initialization. The OMCI protocol is asynchronous: the OLT controller is the master and the ONU controller is the slave. A single OLT controller using multiple protocol instances over separate control channels can control multiple ONUs.

The OLT manages an ONU using OMCI in the following aspects:

- Configuration management: Controls and identifies an ONU, and collects data from and provides data to the ONU.
- Fault management: Supports limited fault management. Most of the operations are limited to failure indication.
- Performance management: Collects and queries performance statistics.
- Security management: Enables/Disables downstream encryption.

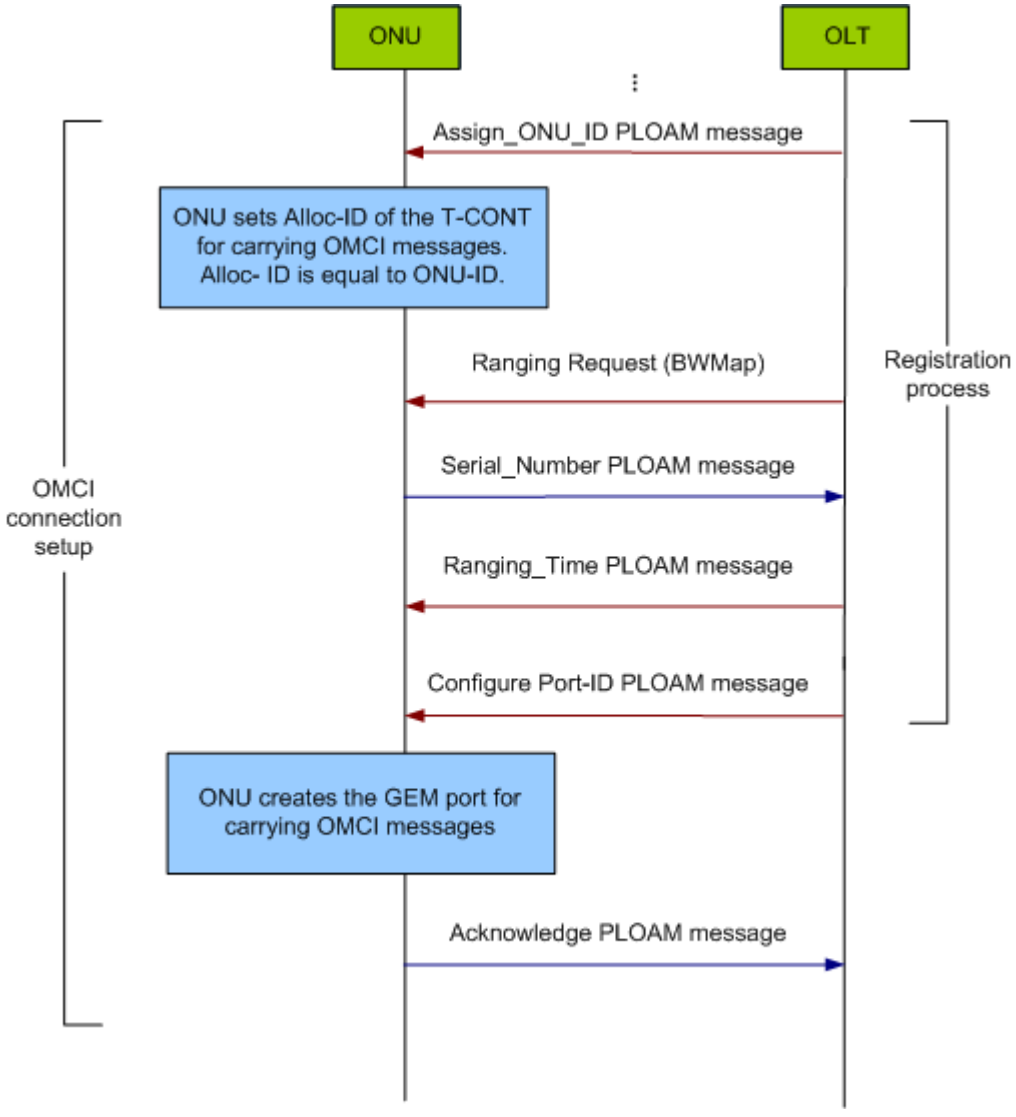
Establishing the ONU Management and Control Channel (OMCC)

Upon initialization, the ONU creates a virtual OMCI T-CONT, an OMCC structure that contains an OMCI queue, a placeholder for an alloc-ID attribute, and – going beyond the definition of an ordinary T-CONT – a placeholder for an OMCI port-ID

 Copyright © Huawei Technologies Co., Ltd.

attribute. The establishment of the OMCC follows the process shown in **Figure 2-15**. During activation, the ONU receives a PLOAM message from the OLT indicating the assignment of the ONU-ID. The ONU populates the alloc-ID attribute of its virtual OMCI T-CONT with the ONU-ID. This makes the alloc-ID for OMCI the same as the assigned ONU-ID. It is therefore not necessary for the OLT to send an assign_alloc-ID message to establish the OMCC. If the OLT nevertheless chooses to send an assign_alloc-ID PLOAM with the default alloc-ID, the ONU should acknowledge this message without taking any specific further action. This is true regardless of the alloc-ID type value in the assign_alloc-ID message: it should not be possible to de-allocate the default alloc-ID with an assign_alloc-ID type 255 message.

Figure 2-15 OMCC establishment



Upon completion of ONU activation in G.984 systems, the OLT assigns a GEM port-ID to the ONU for OMCI messages. This is accomplished by a configure_port-ID PLOAM message. The ONU populates the OMCI port-ID attribute of the OMCC structure based on that message, and responds back to the OLT with an acknowledgment.

 Copyright © Huawei Technologies Co., Ltd.

In G.987 systems, the GEM port for OMCI use is automatically assigned, and is equal to the ONU-ID.

At this point, the OMCC path has been successfully established.

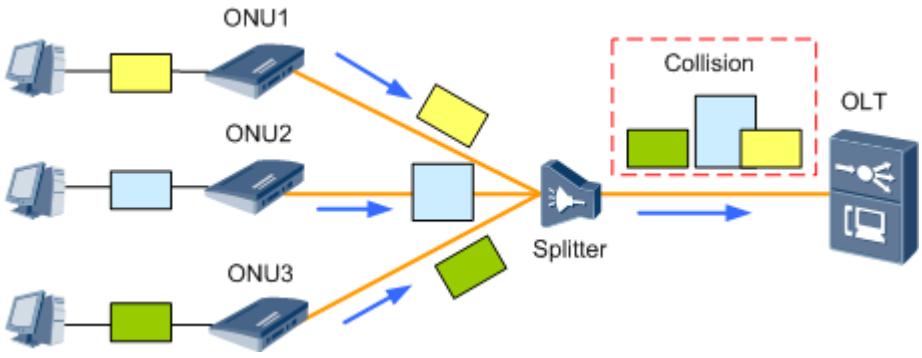
2.4 Key GPON Techniques

2.4.1 GPON Ranging

Why Is Ranging Required

The logic reaches from ONUs to an OLT vary. Therefore, the time required for transmitting optical signals over optical fibers is different and the times when the ONUs receive optical signals is different. In addition, the round trip delays (RTDs) between an OLT and ONUs also vary depending on time and environment. Therefore, collisions may occur when ONU sends data in TDMA mode (in this mode, only one of the ONUs connecting to a PON port sends data at a moment), as shown in **Figure 2-16**. The OLT must precisely measure the distances between itself and each ONU to provide a proper timeslot for converged upstream data from all ONUs to prevent data conflict. In this way, the OLT controls the time for each ONU to send data upstream.

Figure 2-16 Cell transmission without ranging



Ranging Principles

Ranging process is as follows:

- The OLT starts ranging for an ONU when the ONU registers with the OLT for the first time and obtains the round trip delay (RTD) of the ONU. Based on the RTD, the OLT calculates the physical reach of this ONU.
- The OLT specifies a proper equalization delay (EqD) for the ONU based on the physical reach.

NOTE

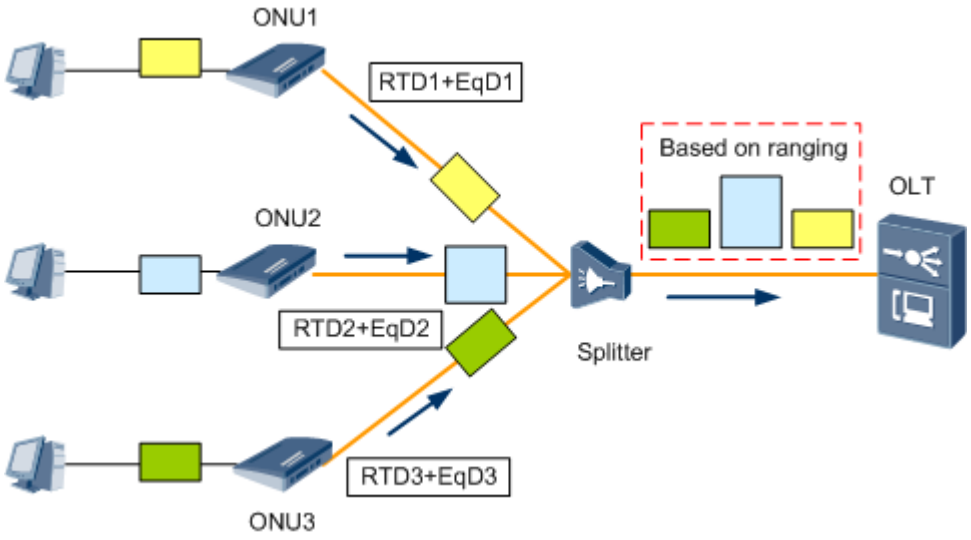
The OLT requires a quiet zone during ranging to pause the upstream transmission channel of the ONUs connected to it. The quiet zone is implemented by emptying BWmap so that no timeslot is allocated for data transmission.

 Copyright © Huawei Technologies Co., Ltd.

Ranging Results

RTD and EqD synchronize data frames sent by all ONUs, preventing data conflict on optical splitters. In this way, all ONUs locate at the same logic reach and they send data at specified timeslots, thereby preventing upstream cell conflict.

Figure 2-17 Cell transmission with ranging

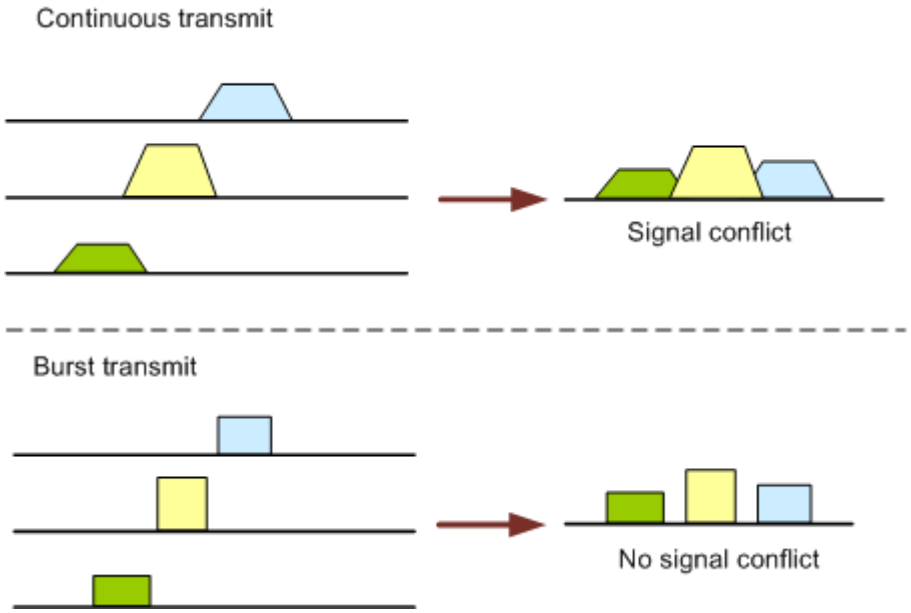


2.4.2 GPON Burst Optical/Electrical Technology

TDMA is used in GPON upstream direction. An ONU transmits data only within the allocated timeslots. In the timeslots that are not allocated to it, the ONU immediately disables the transmission of its optical transceiver to prevent other ONUs from being affected. The OLT then receives the upstream data from each ONU in a burst manner based on timeslots. Therefore, both OLT and ONU optical modules must support burst receive and transmit function to ensure normal running of the GPON system. **Figure 2-18** shows the burst transmit function supported by ONU optical modules, and **Figure 2-19** shows the burst receive function supported by OLT optical modules.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-18 Burst transmit function supported by ONU optical modules

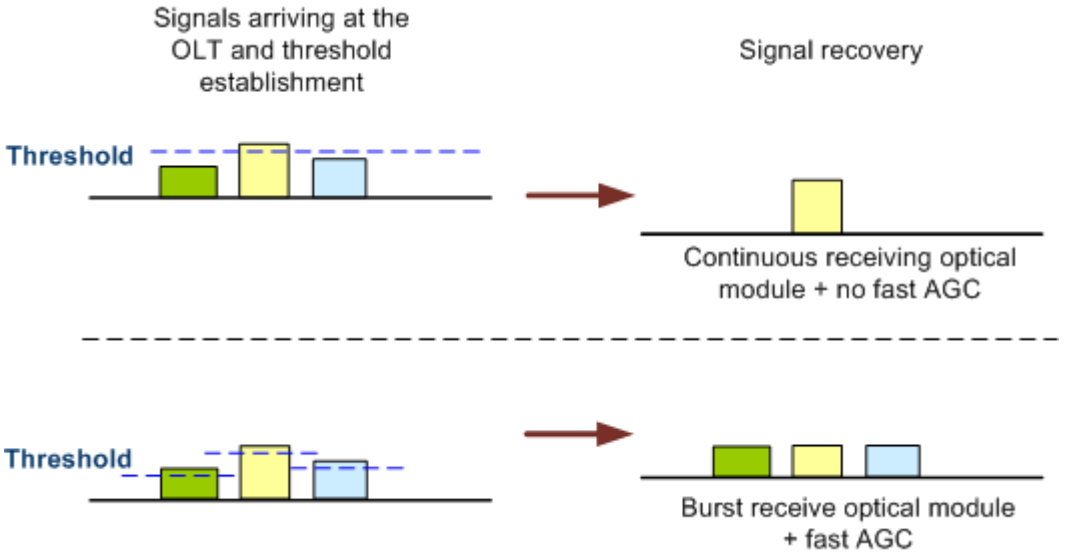


Ranging can be implemented to prevent cells transmitted by different ONUs from conflicting with each other on the OLT. However, the ranging accuracy is ± 1 bit and the cells transmitted by different ONUs have a protection time of several bits (not a multiple of 1 bit). If the ONU optical modules do not support the burst receive and transmit function, the transmitted signals overlap and distortion occurs.

NOTE

In the GPON system, all data is broadcast downstream to ONUs. The transmission requires OLT optical modules to transmit optical signals continuously and ONU optical modules to receive optical signals continuously. Therefore, in GPON downstream direction, the OLT optical modules are not required to support the burst transmit function, the ONU optical modules are not required to support the burst receive function.

Figure 2-19 Burst receive function supported by OLT optical modules



 Copyright © Huawei Technologies Co., Ltd.

- The distance from each ONU to the OLT varies and therefore the optical signal attenuation varies for each ONU. As a result, the power and level of packets received by an OLT at different timeslots various.
- If the OLT-side optical modules do not support the burst receive function, the OLT may restore incorrect signals because only the level greater than the threshold is considered valid and the signals with the level lower than the threshold cannot be restored. The automatic gain control (AGC) function dynamically adjusts the threshold of the receive optical power on the OLT based on the strength of the received optical signals to ensure that the signals of all ONUs can be restored.

2.4.3 GPON DBA

In the GPON system, the OLT controls an ONU's upstream data traffic by sending authorization signals to the ONU. PON requires an effective TDMA mechanism to control the upstream traffic so that data packets from multiple ONUs do not collide in upstream transmission. However, the collision-based mechanism needs to manage QoS in the passive ODN of the PON, which is physically impossible or causes severe efficiency deterioration.

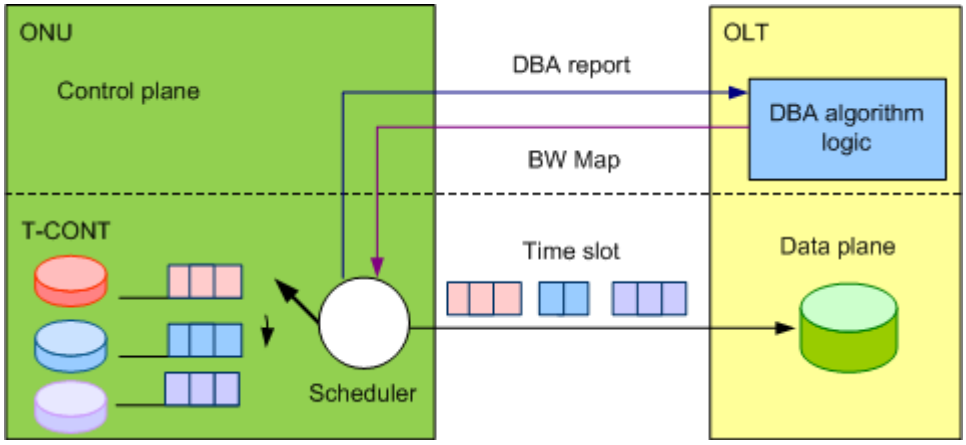
A mechanism for upstream GPON traffic management has been a primary focus in standardization of GPON traffic management. To resolve the problem, ITU-T Recommendation G.984.3 is developed, which defines the DBA protocol for managing upstream PON traffic.

DBA enables the OLT to monitor congestion on the PON network in real time. Then, the OLT can dynamically adjust bandwidths based on congestion, bandwidth usages, and configurations. DBA supports the following functions:

- Improves upstream bandwidth usages on a PON port.
- Supports more users on a PON port.
- Provides higher bandwidths for users, especially the services with significant bandwidth bursts.

Figure 2-20 shows DBA principles.

Figure 2-20 DBA principles



 Copyright © Huawei Technologies Co., Ltd.

- The embedded DBA module of an OLT continuously collects DBA reports, performs calculation, and uses the BWMap field in the downstream frame to notify the ONU of the DBA calculation result.
- According to the BWMap information, the ONUs send data upstream in the timeslots allocated to them, and occupy the upstream bandwidth. Therefore, each ONU dynamically adjusts its upstream bandwidth according to its actually transmitted data traffic, improving upstream bandwidth usage.

Bandwidth can also be allocated in static mode, or fixed mode. In this mode, an OLT periodically allocates a fixed bandwidth to each ONU based on the ONU's service level agreement (SLA), bandwidth, and delay indicators.

- In fixed mode, an OLT uses a polling mechanism. The bandwidths allocated to ONUs may vary but the bandwidth allocated to each ONU is the same in each polling period. The bandwidth guarantee depends on an ONU's SLA but not on its upstream service traffic. An ONU is allocated a fixed bandwidth even carrying no upstream services.
- The static bandwidth allocation method is simple and easy to implement, and is suitable for carrying services with fixed traffic such as TDM. However, the upstream bandwidth cannot be adjusted in real time according to the traffic on the ONU, and the bandwidth utilization is low when the IP service with high burst traffic is carried.

2.4.4 GPON FEC

In actual applications, the transmission of digital signals introduces bit errors and jitter, which degrade signal transmission quality.

To resolve the preceding issue, an error correction technology is required. Among the error correction technologies, the effective ones achieve transmission reliability by reducing bandwidth usages, which also increases telecom device complexity. The error correction technologies are used for controlling errors. The codes involved in these technologies are classified as error detection codes and error correction codes based on usage scenarios.

- Error detection codes, such as parity check codes, are used for detecting error codes.
- Error correction codes, such as BCH codes, Reed-Solomon (RS) codes, and Hamming codes, are used for automatically correcting errors.

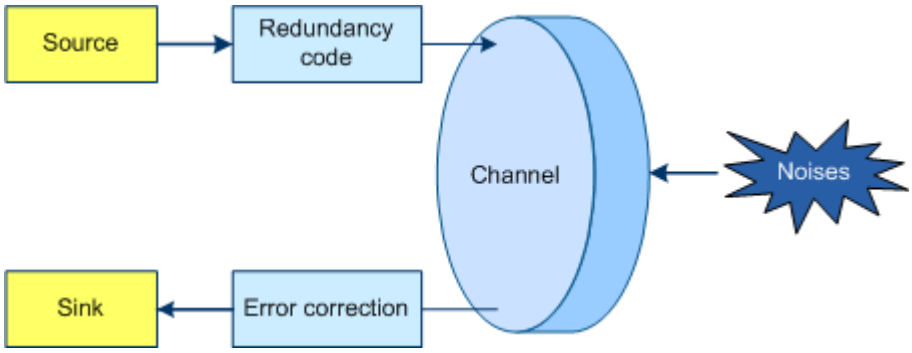
The only difference between the error detection codes and error correction codes lies in performance parameters applied in different usage scenarios. FEC uses error correction codes.

FEC is a data coding technology, which enables the RX end to check error bits in transmission based on the coding data. FEC is unidirectional, not supporting error information feedback.

Redundant codes are added to signals on the TX end. Then, the RX end checks the signals for errors based on error-correcting code (ECC) and corrects errors is there is any. Common FEC codes include Hamming codes, RS codes, and convolutional codes. **Figure 2-21** shows FEC principles.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-21 FEC principles



In the GPON FEC algorithm, the most common RS code RS (255,239) is used, where the code word is 255 bytes long, consisting of 239 data bytes followed by 16 overhead redundant bytes. However, due to the overhead caused by multi-frame tail fragments, the bandwidth throughput of the GPON system with FEC enabled is about 90% of that with FEC disabled. RS code RS (255,239) complies with ITU-T Recommendation G.984.3. The FEC algorithm decreases the bit error rate (BER) from 10-3 to 10-12 on GPON lines.

FEC characteristics are as follows:

- Does not require data retransmission, thereby improving real-time efficiency.
- Enables lines to tolerate louder noises on a basis of a higher bandwidth overhead. (In this case, users must balance between the transmission quality and the bandwidth usage based on site requirements.)
- Based on the preceding characteristics, FEC applies to: The services requiring error detection and correction at the RX end without retransmission.
- Data transmission if the network is in a poor condition. For example, the transmission distance from the OLT to an ONT is long or the transmission line is of poor quality, which results in insufficient optical power budget or high BERs.
- The services requiring no delays (a retransmission prolongs the delay).
- FEC status can be configured in GPON systems based on GPON ports in the downstream direction (by running the **port fec** command) and based on ONUs in the upstream direction. To configure the FEC status in the upstream direction based on ONUs, run either of the following commands:
  - In profile mode, run the **fec-upstream** command.
  - In discrete mode, run the **ont fec-upstream** command.

2.4.5 GPON Line Encryption

In a GPON system, downstream data is broadcast to all ONUs. Then, unauthorized ONUs can receive the downstream data of authorized ONUs, causing system risks.

Line encryption is used to eliminate these security risks. The GPON system uses the Advanced Encryption Standard 128 (AES128) algorithm to encrypt the data packets transmitted in plaintext mode so that the packets are transmitted in ciphertext mode, improving system security. Enable line encryption if the usage scenarios promote high security requirements.

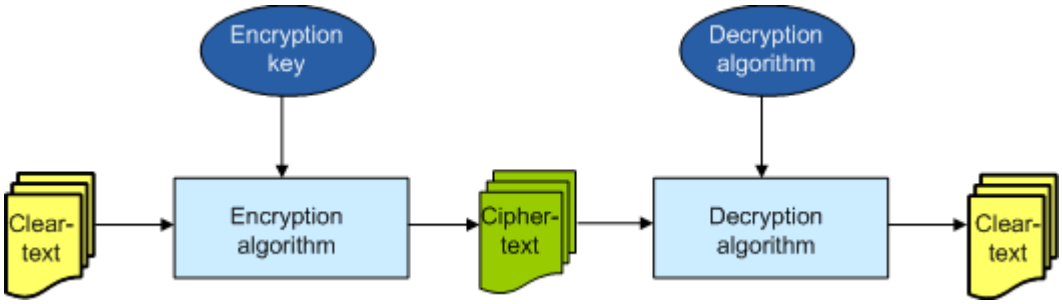
- The line encryption algorithms used in GPON systems neither increase overhead nor decrease bandwidth usages.

 Copyright © Huawei Technologies Co., Ltd.

- The line encryption algorithms will not prolong transmission delays.

Figure 2-22 shows line encryption process.

Figure 2-22 Line encryption process



Key Exchange and Switchover

1. The OLT initiates a key exchange request to the ONU. The ONU responds to the request and sends a new key to the OLT.
2. After receiving the new key, the OLT switches the key to the new one and uses the new key to encrypt data.
3. The OLT sends the frame number that uses the new key to the ONU.
4. The ONU receives the frame number and switches the verification key on data frames.

NOTE

- Due to length limitation on PLOAM messages, the ONU sends the key to the OLT in two pieces and sends both parts of the key three times for extra redundancy. If the OLT is unsuccessful in receiving either part of the key all three times it is transmitted, the OLT initiates a key exchange request to the ONU again until the OLT receives the same key for three times.
- The OLT issues a command three times to the ONU to notify the ONU of using the frame number of the new key. The ONU switches the verification key on data frames after receiving the command only once.

Configuration Method

- In GPON systems, run either of the following commands to configure line encryption status based on GEM ports (excluding multicast and broadcast GEM ports).
  - In profile mode, run the **gem add** command.
  - In discrete mode, run the **gemport add** command.
- Run either of the following commands to encrypt a GEM port in ONT management and control channels (OMCCs):
  - In profile mode, run the **omcc encrypt** command.
  - In discrete mode, run the **ont omcc encrypt** command.

2.4.6 GPON Energy Conservation

Energy conversion enables the OLT to periodically shut down an ONU optical module when the ONU is idle, thereby conserving energy for GPON lines.

 Copyright © Huawei Technologies Co., Ltd.

Overview

An ONU optical module is still working when the ONU is idle. Idle indicates that the traffic within the detection period is less than the specified threshold. In such a case, the OLT can periodically shut down the ONU optical module so that it does not transmit or receive data anymore. This configuration reduces ONU power consumption and conserves energy.

NOTE

Energy conservation is recommended for the FTTH scenario, which is more effective than in other scenarios.

Principles

The energy conservation feature supports the doze mode and cyclic sleep mode, which comply with the ITU-T G987.3 and G.984.3 recommendations.

- **Doze Implementation**
  - After an ONU enters doze mode, the OLT shuts down the ONU optical module in the TX direction. In such a case, the ONU can only receive downstream data from the OLT.
  - In doze mode, if there is traffic on the ONU waiting to be transmitted upstream, the ONU can end the shutdown state by using a local event, so that the transmit function of the ONU optical module is restored from the shutdown state to the normal state.
  - In doze mode, if the OLT wants to make the ONU exit the shutdown state (for example, when the OLT needs to upgrade the ONU), the OLT can send an event to wake up the ONU, so that the transmit function of the ONU optical module is restored from the shutdown state to the normal state.
- **Cyclic Sleep Implementation**
  - After an ONU enters the cyclic sleep mode, the OLT shuts down the receive and transmit channels of the ONU optical module. The ONU cannot receive downstream data from the OLT and cannot send upstream data to the OLT.
  - In cyclic sleep mode, if there is traffic on the ONU waiting to be transmitted upstream, the ONU can end the shutdown state by using a local event, so that the transmit channel of the ONU optical module is restored from the shutdown state to the normal state.
  - In cyclic sleep mode, if the OLT wants to make the ONU exit the shutdown state, the OLT needs to wait till the ONU sleep timer expires and the ONU enters the short wake-up state. The ONU responds to the wake-up event sent by the OLT only after entering this state, and the transmit and receive channel of the ONU optical module will be restored from the shutdown state to the normal state.

Key Techniques

After an energy conservation profile is bound to an ONU, the ONU automatically enters or exits the energy conservation mode if the in-mode or out-of-mode conditions are met.

 Copyright © Huawei Technologies Co., Ltd.

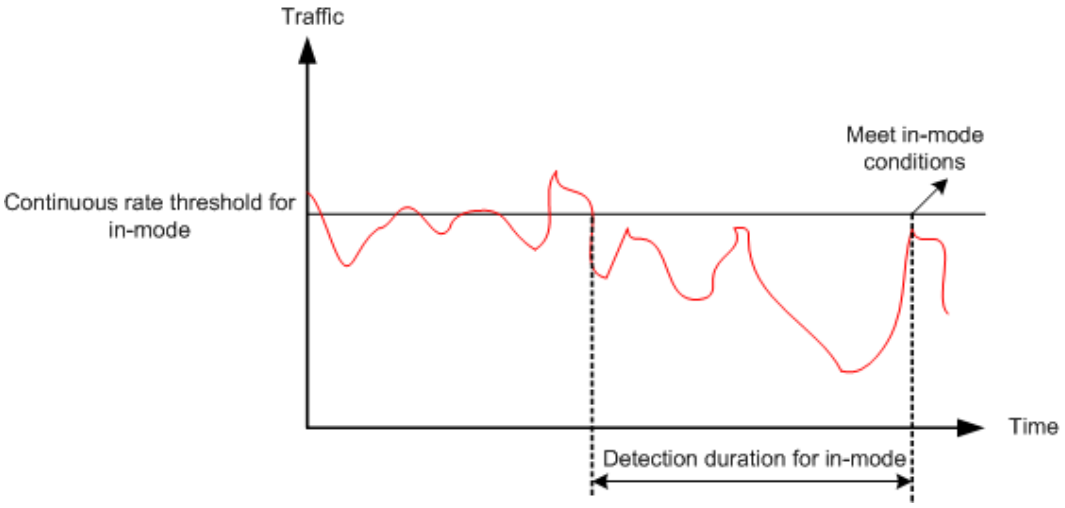
NOTE

In doze mode, the OLT shuts down only the data transmit channel of the ONU. Therefore, the ONU can enter or exit the energy conservation mode as long as its upstream traffic meets the in-mode or out-of-mode condition.

In-Mode Conditions

The energy conservation profile defines the upstream and downstream continuous rate thresholds for ONU in-mode. Within the detection duration, if the upstream and downstream traffic of an ONU is smaller than the configured continuous rate thresholds, the ONU enters the energy conservation mode, as shown in **Figure 2-23**.

Figure 2-23 In-mode conditions



Out-of-Mode Conditions

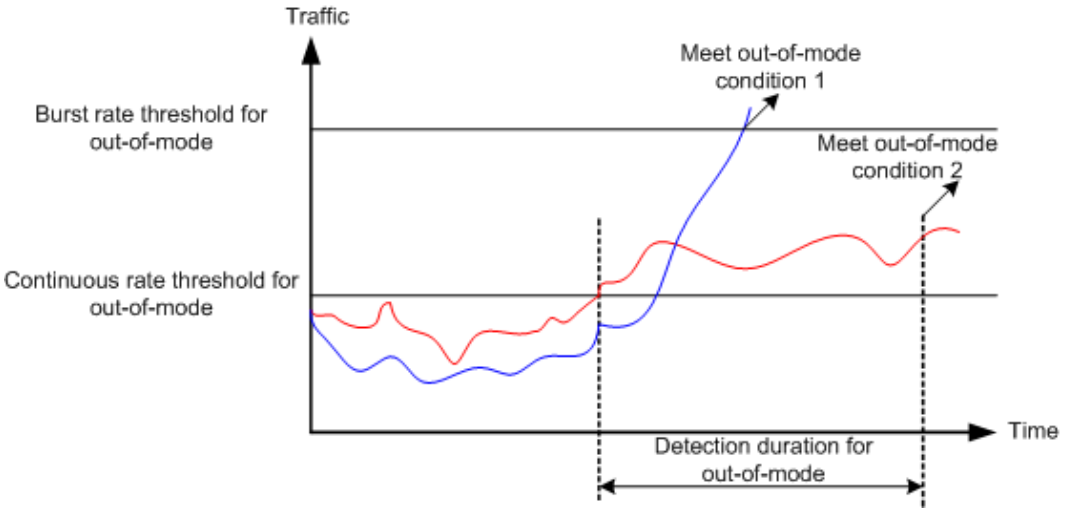
The energy conservation profile defines the upstream and downstream continuous rate thresholds and burst rate thresholds for ONU out-of-mode. An ONU exits the energy conservation mode when its upstream and downstream traffic meets either of the following conditions:

- Out-of-mode condition 1: Within the detection duration, the ONU upstream or downstream burst traffic is larger than the configured burst rate threshold.
- Out-of-mode condition 2: Within the detection duration, the upstream traffic and downstream traffic are larger than the configured continuous rate thresholds.

Figure 2-24 shows out-of-mode conditions.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-24 Out-of-mode conditions



Configuring Energy Conservation

1. Create a GPON ONT energy conservation profile.

   Run the **ont power-reduction-profile add** command to create a GPON ONT energy conservation profile.

   NOTE

   An OLT supports up to 32 ONT energy conservation profiles.

2. Query the information about an energy conservation profile.

   Run the **display ont power-reduction-profile** command to query the information about the energy conservation profile.

3. Bind the energy conservation profile.

   Run the **ont power-reduction-config** command to bind the energy conservation profile to the ONUs connected to a GPON port. After the binding, energy conservation configurations are automatically issued to the ONTs.

   NOTE

   - ONU energy conservation is incompatible with Type B single homing, Type B dual homing, Type C single homing, and Type C dual homing. When both ONU energy conservation and Type X homing are configured, ONU energy conservation fails to take effect.
   - Energy conservation takes effect only between Huawei OLTs and ONUs but not with third-party ONUs.
   - For a GPON port, the ONUs connected to it support the binding of up to 8 energy conservation profiles. For an XG-PON port, the number is 16.

2.4.7 ONU Automatic Registration and Service Flow Creation

 Copyright © Huawei Technologies Co., Ltd.

Overview

The OLT supports the pre-configuration of the service flow creation policy on the PON port. When the ONU that meets conditions goes online, registration and service flow creation are automatically implemented based on the pre-configured policy. This function simplifies the configuration process and improves the installation and deployment efficiency.

NOTE

The ONU ID is configured during the pre-configuration of the service flow creation policy. When the ONU starts to register with the OLT, the ONU ID and vendor identity are verified. Only the ONU meeting conditions can implement automatic registration.

O&M for these automatically created service flows is the same as that for commonly created service flows.

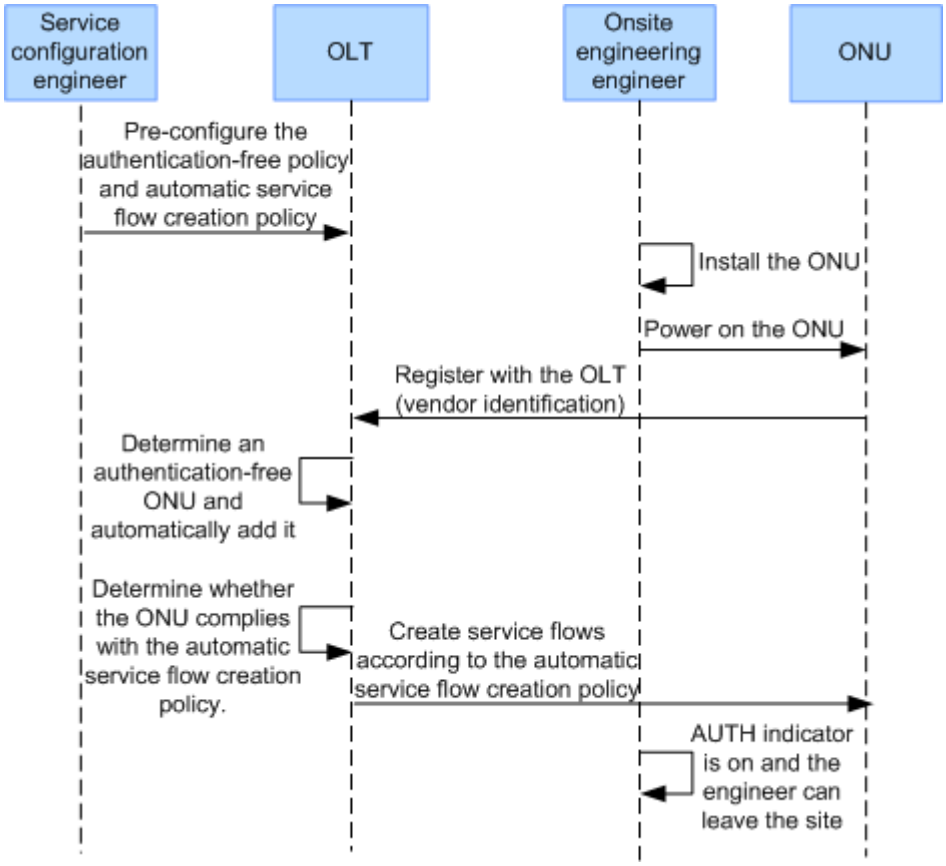
NOTE

Automatic registration and service flow creation are supported in the PON profile mode only. They are not supported in the distributed mode.

Automatic Service Flow Creation

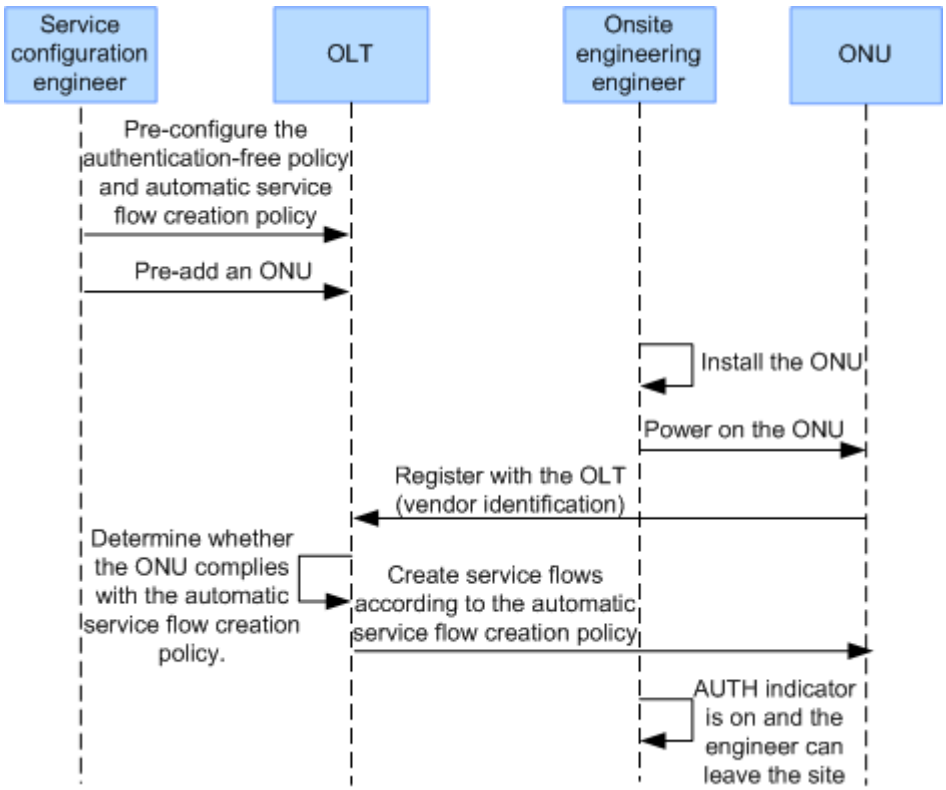
The automatic service flow creation process covers 2 scenarios: no ONU is manually added; the ONU is manually added.

Figure 2-25 Process of automatic service flow creation (no ONU is manually added)



 Copyright © Huawei Technologies Co., Ltd.

Figure 2-26 Process of automatic service flow creation (the ONU is manually added)



Configuring Automatic Service Flow Creation

1. Configure an automatic ONU addition policy.

   Run the **ont auto-add-policy** command to configure an automatic ONU addition policy.

   NOTE

   When **auto-confirm** is set to **enable**, you do not need to run the **ont confirm** command on the GPON port to confirm the ONU. When **auto-confirm** is set to **disable**, you need to run the **ont confirm** command on the GPON port.

2. Configure the automatic service flow creation policy.

   Run the **auto-service-port**

   command to configure the automatic service flow creation policy.

For details about how to configure automatic service flow creation, see "**10.3.3.5 Configuring Automatic Service Flow Creation (GPON)**".

2.4.8 ONU Automatic Migration and Aging

Automatic Migration

When the ONU automatic migration function is enabled, if a PON port or fiber is faulty, the ONU connected to this faulty PON port is migrated to another PON port. If the ONU automatic adding policy is met, the ONU is automatically removed from the original PON port, and then added to the new PON port. In this

 Copyright © Huawei Technologies Co., Ltd.

process, the OLT does not report the SN authentication conflict alarm. The ONU connected to the original PON port must meet the following conditions:

- The ONU is offline.
- The ONU supports OMCI management.
- The ONU is not a member of the type C protection group.
- The ONU supports SN authentication.

NOTE

Automatic migration and aging are supported in the PON profile mode only. They are not supported in the disperse mode.

Automatic Aging

When an offline ONU is aged and meets automatic adding policy, the OLT automatically deletes the aged ONU to save resources. The ONU automatic aging period can be configured. The value range is 1–365. The default value is no-aging, which indicates no aging. For an aged ONU that meets the automatic adding policy, it will be automatically registered after being powered on.

Configuring Automatic Migration and Aging

1. Configure ONU automatic migration.

   Run the **ont auto-migration** command to configure ONU automatic migration.

2. Configure ONU automatic aging.

   Run the **ont auto-aging** command to configure ONU automatic aging.

2.5 GPON Terminal Authentication and Management

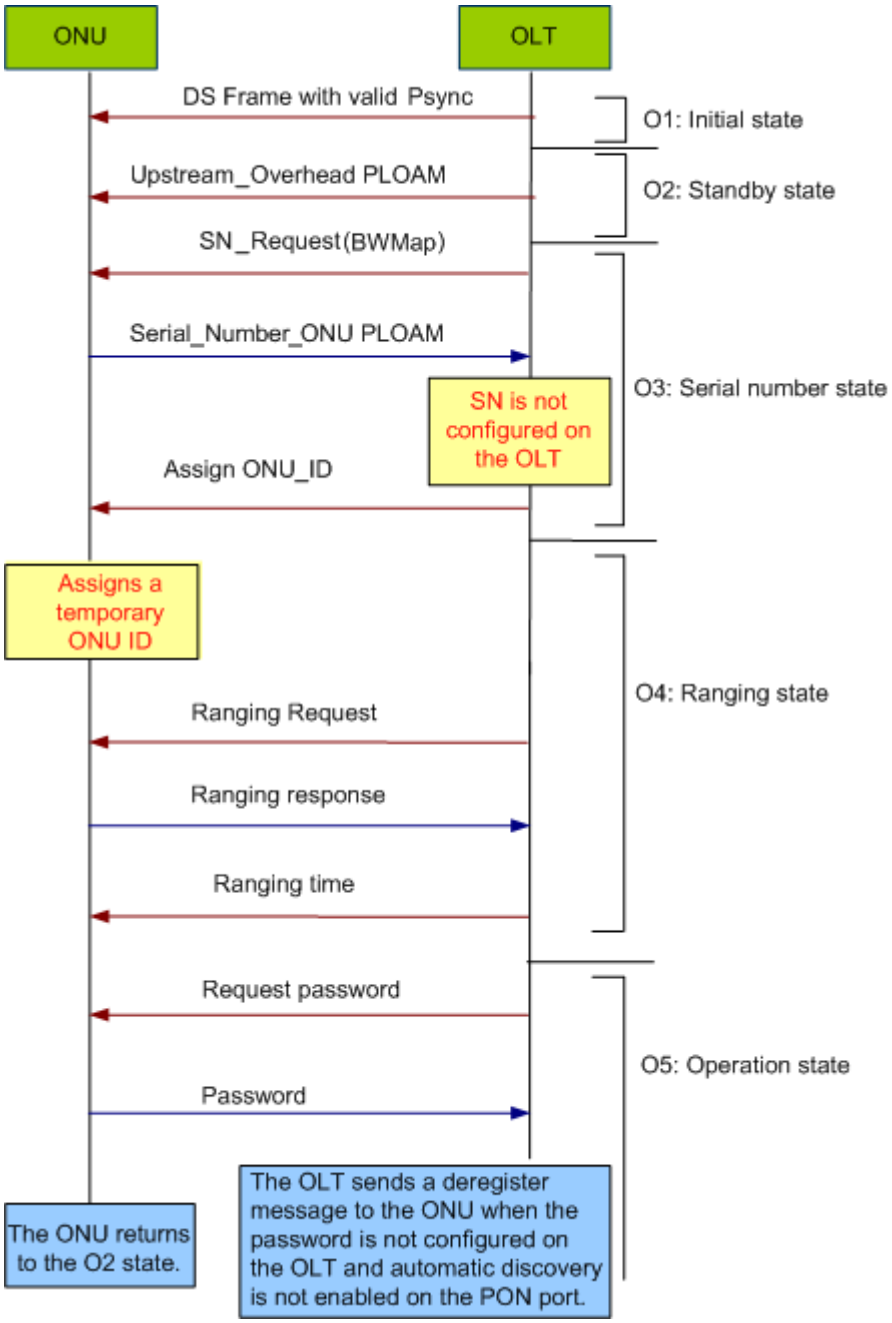
GPON terminal authentication is a mechanism in which an OLT authenticates an ONU according to the authentication information reported by the ONU and in this way denies access to unauthorized ONUs. In the GPON system, only authenticated ONUs can access the system. After the ONU passes authentication and goes online, data can be transmitted between ONUs and the OLT.

2.5.1 GPON Terminal Authentication (ONU Is Not Preconfigured)

The authentication modes for ONUs pre-configured on the OLT include SN, SN +Password, and Password. **Figure 2-27** shows the registration flow for ONUs not pre-configured on the OLT.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-27 Authentication process of an ONU that is not preconfigured



- The OLT sends a serial number (SN) request to the ONU.
- The ONU responds to the SN request message sent from the OLT.
- Upon receiving the SN response from the ONU, the OLT assigns a temporary ONU ID to the ONU.
- After the ONU enters the operation state, the OLT sends a password request message to the ONU. The ONU then responds with a password. The password is not configured on the OLT.
  - If the automatic discovery function is not enabled on the PON port to which the ONU is connected, the OLT sends a deregister message to the

 Copyright © Huawei Technologies Co., Ltd.

ONU. Upon receiving this message, the ONU sends a register request message to the OLT.

– If the automatic discovery function is enabled on the PON port to which the ONU is connected, the port reports an alarm to the command line interface (CLI) or network management system (NMS), indicating that the ONU is automatically discovered. The ONU can go online only after being confirmed.

2.5.2 GPON Terminal Authentication (ONU Has Been Pre-configured)

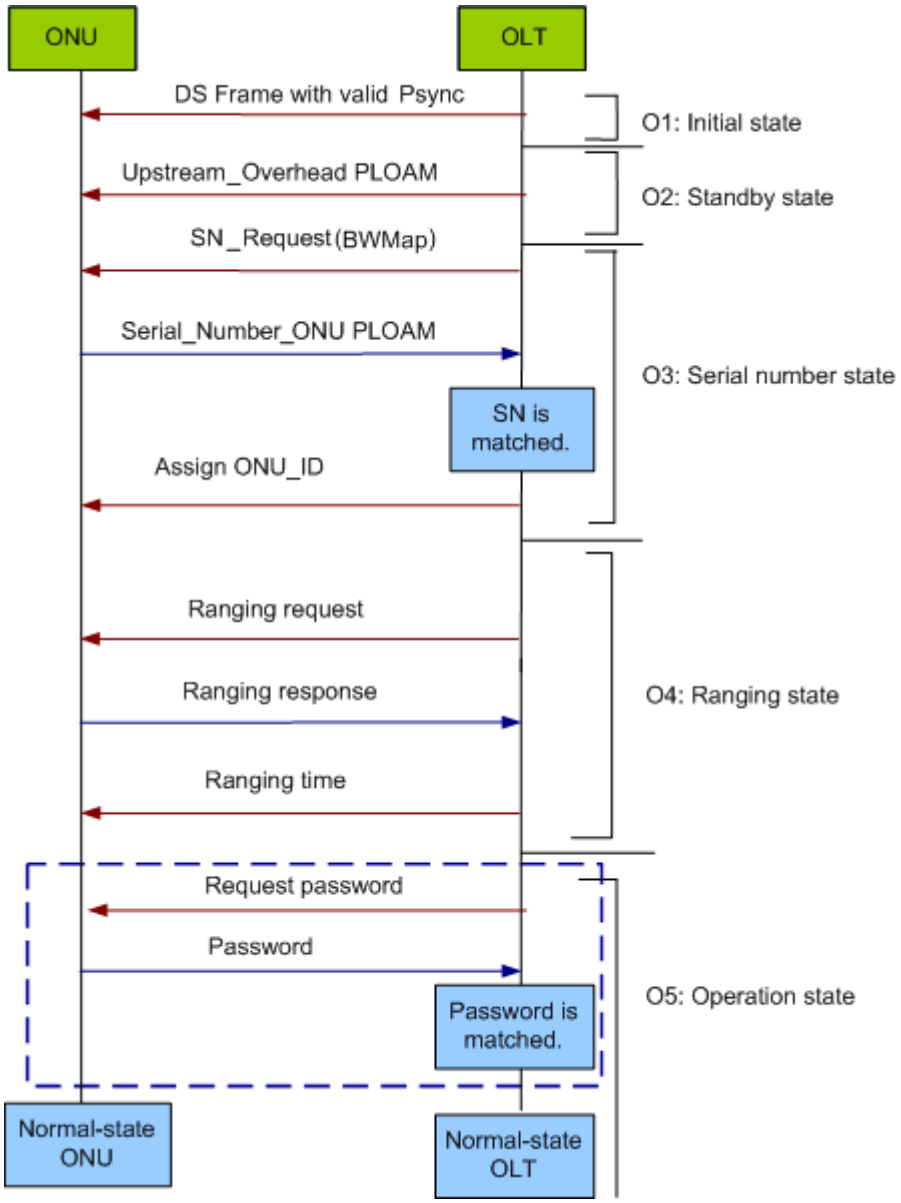
A pre-configured ONU can be authenticated in three modes: SN, SN+password, and password.

SN/SN+Password Authentication

In SN authentication, the OLT matches only the ONU SN. In SN+password authentication, the OLT matches both the ONU SN and password. **Figure 2-28** shows the authentication flow.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-28 SN/SN+password authentication flow



NOTE

If an ONU is authenticated in SN mode, no password is required in the authentication process.

- After receiving an SN response message from an ONU, the OLT checks whether another ONU with the same SN is online. If yes, the OLT reports an SN conflict alarm to the CLI or NMS. If no, the OLT directly assigns a user-defined ONU ID to the ONU.
- After the ONU enters the operation state,
  - For the ONU that is authenticated in SN mode, the OLT does not send a password request message to this ONU. Instead, the OLT automatically configures a GEM port that has the same ID as the ONU ID for the ONU

 Copyright © Huawei Technologies Co., Ltd.

for carrying OMCI messages, and allows the ONU to go online. In addition, the OLT reports an ONU online alarm to the CLI or NMS.

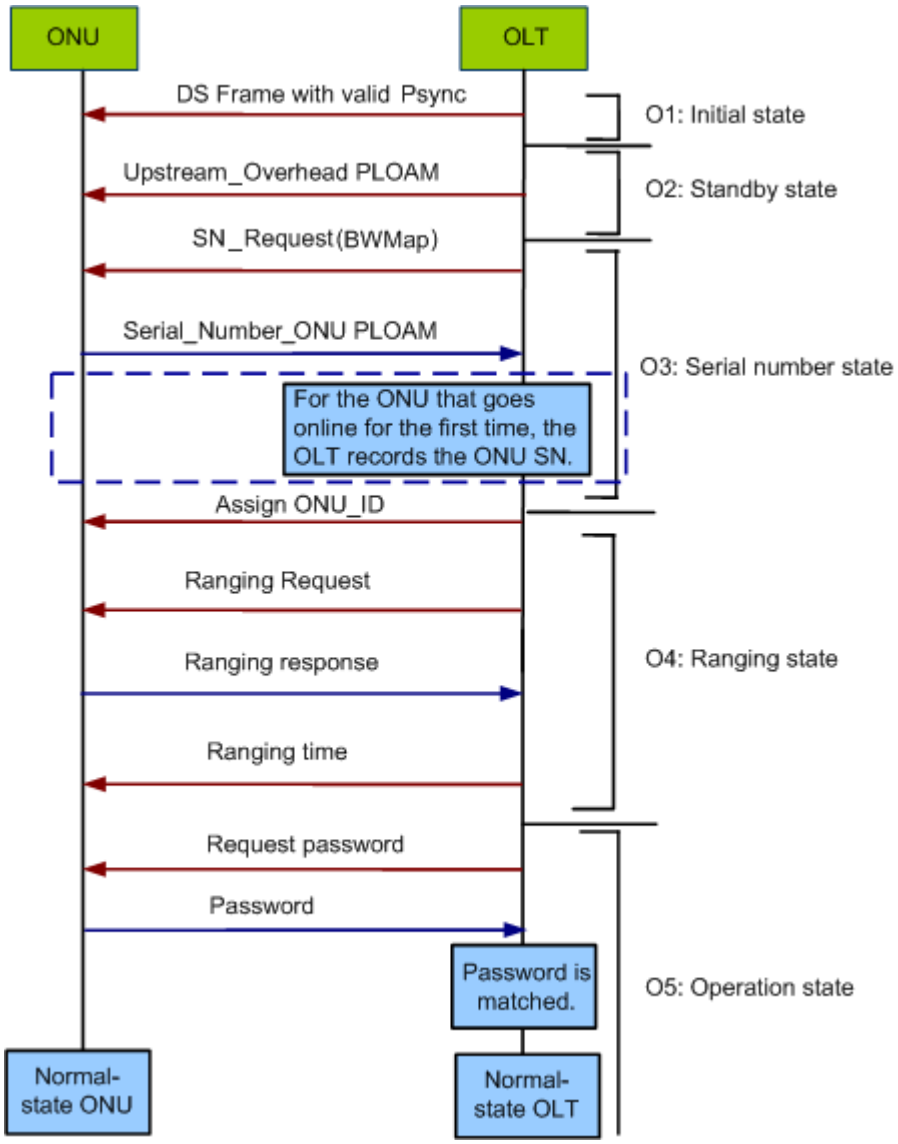
– For the ONU that is authenticated in SN+password mode, the OLT sends a password request to the ONU, and compares the password reported by the ONU with the local password. If the two passwords are the same, the OLT directly configures a GEM port for the ONU to carry OMCI messages, and allows the ONU to go online. In addition, the OLT reports an ONU online alarm to the CLI or NMS. If the two passwords are not the same, the OLT reports a password error alarm to the CLI or NMS. The OLT does not report an ONU automatic discovery message even if the ONU automatic discovery function is enabled on the PON port. Instead, the OLT sends the Deactivate_ONU-ID PLOAM message to deregister the ONU.

Password Authentication

An ONU that uses password authentication is added to a PON port on an OLT in advance, and then this ONU is connected to the PON port. In password authentication, if finding that the SN or password of the ONU to be authenticated conflicts with that of an online ONU, the OLT deregisters the ONU to be authenticated. This does not affect the online ONU. Password authentication is available in two modes: once-on and always-on.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-29 Initial ONU authentication in once-on mode



Once-on Application Scenarios

A carrier allocates a password to a user and requires the user to go online within a specified time. After going online, the user cannot change the ONU. To change the ONU, the user must notify the carrier. In once-on mode, the aging time is configurable. After the aging time is set, the ONU must register with the OLT and go online within the preset aging time. Otherwise, the ONU is not allowed to register with the OLT or go online. Once the ONU is authenticated, its SN cannot be changed.

In once-on mode,

- Only the initial authentication of an ONU is performed by password, as shown in **Figure 2-29**.
- In subsequent authentications, the ONU can be authenticated by SN or SN +password according to the CLI configuration, as shown in **Figure 2-28**.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

In once-on mode, before the ONU registration times out or before the ONU successfully registers with the OLT for the first time, the ONU discovery status is ON. Only the ONU whose discovery status is ON is allowed to register with the OLT and go online. After the ONU registration times out or after the ONU successfully registers with the OLT for the first time, the OLT sets the ONU discovery status to OFF.

- The ONU whose registration times out is not allowed to register with the OLT or go online. The registration timeout flag of the ONU needs to be reset at the central office (CO), and then the ONU can go online.
- An ONU that successfully registers for the first time is allowed to register and go online again.

Always-on Application Scenarios

NOTE

In the Always-on mode, the OLT does not need to record the SN when an ONU goes online for the first time.

The always-on mode applies to the following scenario: A carrier allocates a password to a user, and the user can use different ONUs with this password and different SNs. The user can change the ONU without informing the carrier. In always-on mode, there is no restriction on the time when the user goes online.

- An ONU is authenticated by password when it goes online for the first time. After the ONU passes the password authentication and goes online successfully, the OLT generates an SN+password entry according to the SN and password of the ONU. **Figure 2-29** shows the authentication process.
- The following scenarios are involved if it is not the first time that an ONU goes online:
  - If the SN and password of the ONU are the same as the SN and password of the ONU that successfully goes online for the first time, the ONU is authenticated by SN+password. **Figure 2-28** shows the authentication process.
  - If the user replaces the ONU with an ONU that has the same password but a different SN, the new ONU is authenticated by password. After this ONU passes authentication and goes online successfully, the original SN +password entry is updated. **Figure 2-29** shows the authentication process.

2.5.3 GPON Terminal Management

The ONUs (including MDUs and ONTs) in a GPON system are managed using physical layer OAM (PLOAM) messages and OMCI messages.

PLOAM, defined in ITU-T Recommendation G.984.3, is used for exchanging management and maintenance messages, such as DBA and DBRu messages, between the GPON physical layer and TC layer.

ONUs (including MDUs and ONTs) in a GPON system are managed using OMCI messages. The ONUs support offline deployment, plug-and-play, and automatic service provisioning. For details about OMCI management functions, see **OMCI**.

- OMCI messages are used for maintaining and managing service hierarchies, such as discovering device hardware capabilities and configuring alarm maintenance and service capabilities.

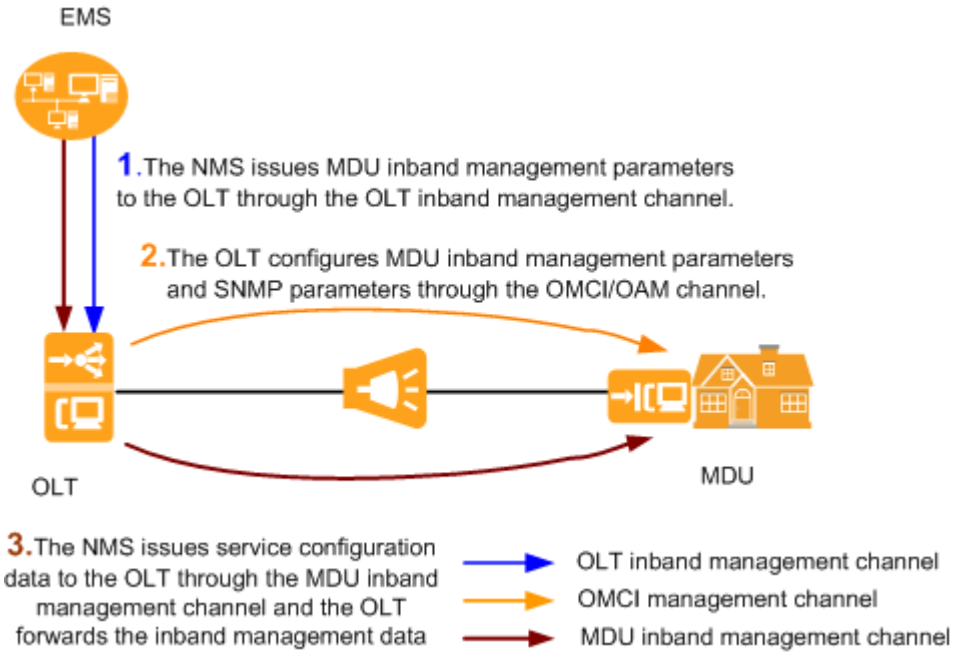
 Copyright © Huawei Technologies Co., Ltd.

- OMCI enables ONUs to support offline configuration so that the ONUs do not need to store configuration data locally, facilitating service provisioning.

MDU Management

Figure 2-30 shows the process of configuring a management channel for an MDU.

Figure 2-30 Process of configuring a management channel for an MDU



1. The NMS issues MDU inband management parameters to the OLT through the OLT inband management channel.
2. The OLT configures the MDU inband management parameters and Simple Network Management Protocol (SNMP) parameters through the OMCI or OAM channel to set up the MDU inband management channel.
3. The NMS issues service configuration data through the MDU inband management channel. After the MDU inband management channel is set up, the NMS configures and manages the MDU through the SNMP channel. In such a manner, the OLT only needs to forward the MDU inband management data.

ONT Management

GPON terminals are managed using one of these protocols: optical network terminal management and control interface (OMCI), Extensible Markup Language (XML), or Technical Report 069 (TR069).

- The optical network terminal management and control interface (OMCI) protocol is defined by ITU-T G984.4, which applies to managing optical network terminals (ONTs) in a GPON system. Huawei ONTs comply with OMCI. OMCI messages are transmitted between an optical line terminal (OLT) and an ONT over a dedicated permanent virtual channel (PVC) in asynchronous transfer mode (ATM) or a GPON encapsulation mode (GEM) port. The OMCI protocol manages and provides O&M for the ONT.

 Copyright © Huawei Technologies Co., Ltd.

- Extensible Markup Language (XML) is a text format used for message interaction between devices. The network management system (NMS) uses XML to manage ONTs in a Huawei FTTx system. XML is also a management mode extended from OAM because not all voice and Layer 3 gateway services are defined in the OAM.
- Technical Report 069 (TR069) is a network management protocol defined by the DSL Forum. The full name of TR069 is CPE WAN Management Protocol (CWMP). CPE is the acronym for customer premises equipment and WAN is the acronym for wide area network. It defines a new network management structure consisting of management models, interaction interfaces, and basic management parameters. In the network management structure, the management server functions as an Auto-Configuration Server (ACS) and is responsible for managing the CPE. The ACS and CPE use Hypertext Transfer Protocol (HTTP) to communicate with each other. The ACS serves as an HTTP server and the CPE serves as a client. Management operations are implemented using XML-based remote procedure call (RPC).

Optical network terminals (ONTs) are classified into three types: bridge type, bridge+voice type, and gateway type.

- A bridge-type ONT provides Layer 2 data and multicast services.
- A bridge+voice-type ONT provides Layer 2 data, Layer 2 multicast services, and voice over IP (VoIP) services.
- A gateway-type ONT provides Layer 3 data, Layer 3 multicast services, and VoIP services.

Each type of terminal management solution has a unique service management scope. Based on terminal types, different GPON terminal management solutions are available, including OMCI+NMS, OMCI+NMS+XML, OMCI+NMS+TR069.

- The OMCI+NMS solution manages Layer 2 services, voice services and the PON link layer. This solution cannot manage Layer 3 services.
- The OMCI+NMS+XML solution manages Layer 2 services, Layer 3 services, and voice services.
- The OMCI+NMS+TR069 solution manages Layer 3 services and voice services, and remotely identifies faults. When this solution is used, OMCI is still required to manage Layer 2 services and the PON link layer.

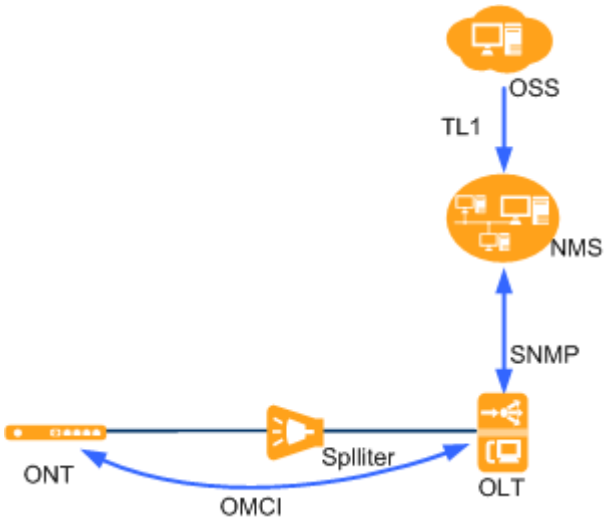
OMCI+NMS

A standard optical network terminal management and control interface (OMCI) solution enables you to manage optical network terminals (ONTs) supplied by different vendors in diverse types of scenarios. An optical line terminal (OLT) and an ONT are closely coupled with each other. If a new service requirement is not defined in the OMCI, a new OMCI entity must be defined. An OMCI solution enables you to manage Layer 2 features and voice services. The OLT communicates with the ONT in OMCI mode.

Figure 2-31 shows the general principles of the OMCI+NMS solution for NMS +OLT+ONT deployment scenarios.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-31 General principles of the OMCI+NMS solution



1. The Operations Support System (OSS) issues service configuration parameters to the NMS using the TL1 northbound interface (NBI).
2. The NMS uses the Simple Network Management Protocol (SNMP) to manage the OLT.
3. The OLT issues service configuration parameters to the ONT through an OMCI channel.

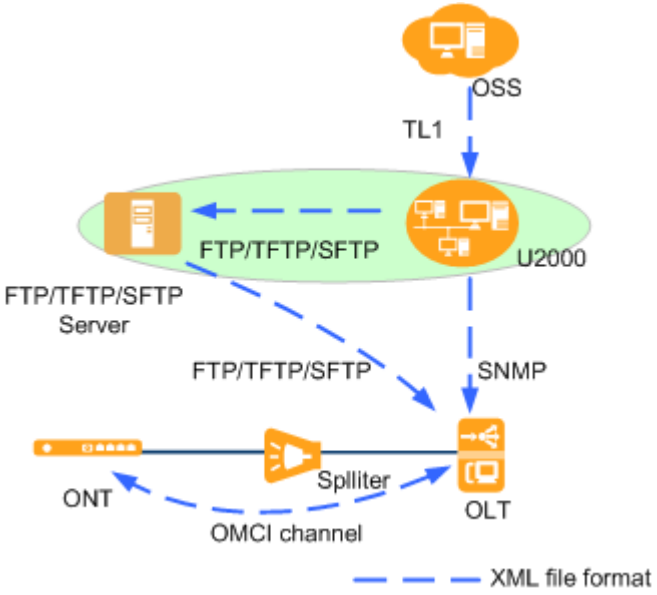
OMCI+NMS+XML

To overcome the limitations of the OMCI+NMS solution, Huawei provides the OMCI+NMS+XML solution. In this solution, the NMS uses XML files transmitted over an IP channel to communicate with the OLT, and the OLT uses XML files transmitted over an OMCI channel to communicate with ONTs. The OMCI protocol manages Layer 2 services and the XML protocol manages Layer 3 and voice services.

Figure 2-32 shows the general principles of the OMCI+NMS+XML solution for NMS+OLT+ONT deployment scenarios.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-32 General principles of the OMCI+NMS+XML solution



The NMS uploads the XML file to the FTP/TFTP/SFTP server. The OLT obtains the XML file from the FTP/TFTP/SFTP server and transparently transmits the XML file to an ONT through the OMCI channel.

NOTE

TFTP/FTP has security risks due to protocol limitations. Therefore, the SFTP mode with high security is recommended for loading the XML file to the ONT.

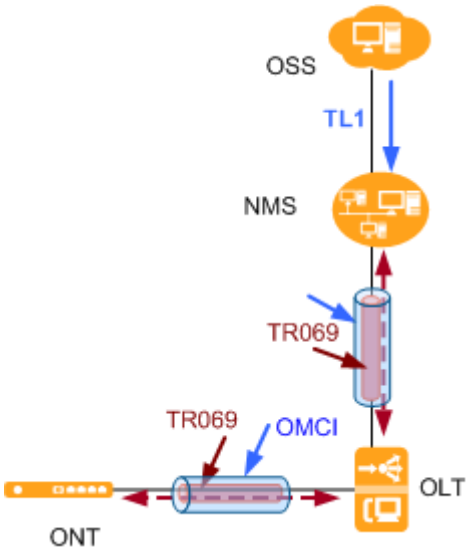
1. The OSS issues service configuration parameters to the NMS using the TL1 NBI.
2. The NMS converts service information to XML files and uploads the files to the FTP/TFTP/SFTP server.
3. The NMS issues ONT configuration update commands to download the files.
4. The OLT obtains the XML files from the FTP/TFTP/SFTP server.
5. The OLT issues the XML files to the ONT through the OMCI channel.
6. The ONT returns execution results to the OLT using the OMCI entity.
7. The OLT reports the results to the NMS using traps.

The OMCI+NMS+XML solution meets all requirements for configuring ONTs but configuration files are transmitted in unidirectional mode. Due to this limitation, the configuration files only implement service configurations and status performance management, but cannot provide operation and maintenance (O&M) functions such as query of ONT status and configuration, and test and diagnose functions. To overcome OMCI+NMS+XML limitations, Huawei provides TR069 over OMCI. As a supplement to OMCI+NMS+XML, TR069 over OMCI is used for remote O&M and fault diagnosis. The NMS can use TR069 to remotely maintain ONTs without a dedicated TR069 server.

Figure 2-33 shows the general principles for NMS+OLT+ONT deployment scenarios.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-33 General principles of the TR069 over OMCI solution



The solution manages configuration, performance, faults, and status of IP-based services by applying the associated methods described in the TR069 solution to the OMCI solution. The OLT and ONT transparently transmit data between each other.

1. The NMS manages and maintains an ONT, and queries the ONT status. It encapsulates involved commands to character strings or binary code flows in a certain format, and issues the commands to the OLT through the MIB interface.
2. The OLT transparently transmits the character strings or binary code streams to the ONT using an extended OMCI entity.
3. The ONT returns execution results to the OLT using the OMCI entity.
4. The OLT reports the results to the NMS using traps.

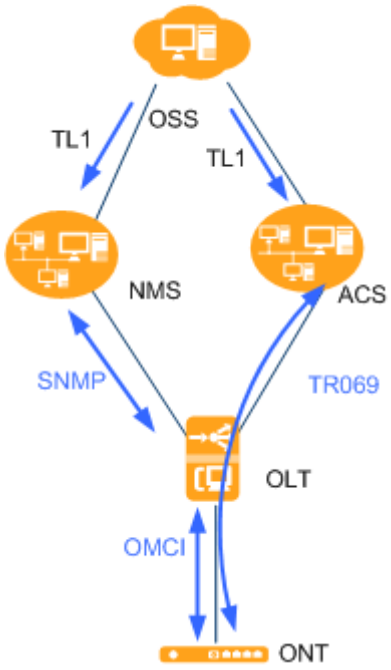
OMCI+NMS+TR069

This solution allows an Auto-Configuration Server (ACS) to manage all the terminals on the network, locate faults, provide services, and collect performance statistics. Based on SNMP and TR069, this solution allows the ACS to manage home terminals in a unified manner, reducing O&M costs. TR069 automatically implements ONT configuration, dynamically provisions services, remotely locates faults, and rapidly collects terminal statistics.

Figure 2-34 shows the general principles of the solution.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-34 General principles of the OMCI+NMS+TR069 solution



This solution allows the NMS to manage the OLT using SNMP, manage voice and Layer 3 services using TR069, and manage PON link layer using OMCI.

1. The OSS issues service configuration parameters to the NMS using the TL1 NBI.
2. The NMS manages the OLT using SNMP.
3. The OLT issues PON link layer configuration to the ONT using OMCI.
4. The ONT returns execution results to the OLT. Then the IP channel is set up.
5. The ONT registers with the ACS.
6. The ACS encapsulates user information in a TR069-compliant format and sends it to the ONT through the IP channel. The user information includes operations, maintenance items, and queries performed by a user. The IP channel is bidirectional.

Advantages and Disadvantages of the Terminal Management Solutions

Bridge type, bridge+voice type, and gateway type ONTs provide different types of services. Therefore, different solutions are used to manage these ONTs. **Table 2-4** lists the advantages and disadvantages of each solution. **Table 2-5** lists the recommended solution for each type of ONT.

 Copyright © Huawei Technologies Co., Ltd.

Table 2-4 Advantages and disadvantages of each solution

Terminal Management Solution	Advantage	Disadvantage
OMCI+NMS	• A unified interface is used for ONT service management. • The OLT and ONT communicate with each other using OMCI-associated standards. • The ONT does not require a management IP address.	• The OLT and ONT are closely coupled with on each other. New services on the ONT require the OLT's support, adding to the difficulty in deploying new services. • The OMCI standard is not fully developed. If a new service requirement is not defined in the OMCI, a new OMCI entity must be defined.
OMCI+NMS+XML	• The ONT does not require a management IP address. • The OLT and ONT are not closely coupled with each other to certain extent. • A unified management server is used for swift service deployment.	• This is a Huawei's proprietary solution and cannot interact with devices from other vendors. • Voice and Layer 3 services cannot be configured using a command on the OLT.

 Copyright © Huawei Technologies Co., Ltd.

Terminal Management Solution	Advantage	Disadvantage
OMCI+NMS+TR069	• An OLT version and an ONT version are not bound to each other. In other words, an OLT upgrade does not require an ONT upgrade; the opposite is also true. • TR069 provides an enhanced definition and deployment scenario for the IP-based customer premises equipment (CPE) service management model. Therefore, ONT vendors can easily deploy new gateway and voice services.	• TR069 is based on the IP protocol and requires an extra IP management network. • Different interfaces are used to manage the ONT. The EMS manages the link layer and the ACS manages IP-based services.

Table 2-5 Recommended solutions for each type of ONT

Terminal Type	Optional Solution	Recommended Solution
Bridge type	OMCI+NMS or OMCI +NMS+XML	OMCI+NMS
Bridge+voice type	OMCI+NMS or OMCI +NMS+XML	OMCI+NMS+XML (NMS provisions services) OMCI+NMS (OLT is connected to third-party ONTs)
Gateway type	OMCI+NMS+XML or OMCI+NMS+TR069	Only the NMS server is deployed: OMCI+NMS +XML The NMS and TR069 servers are deployed: OMCI+NMS+TR069

2.6 Continuous-Mode ONU Detection

This function is used to detect the continuous-mode ONU of terminal devices such as the MDU and ONT.

 Copyright © Huawei Technologies Co., Ltd.

Overview

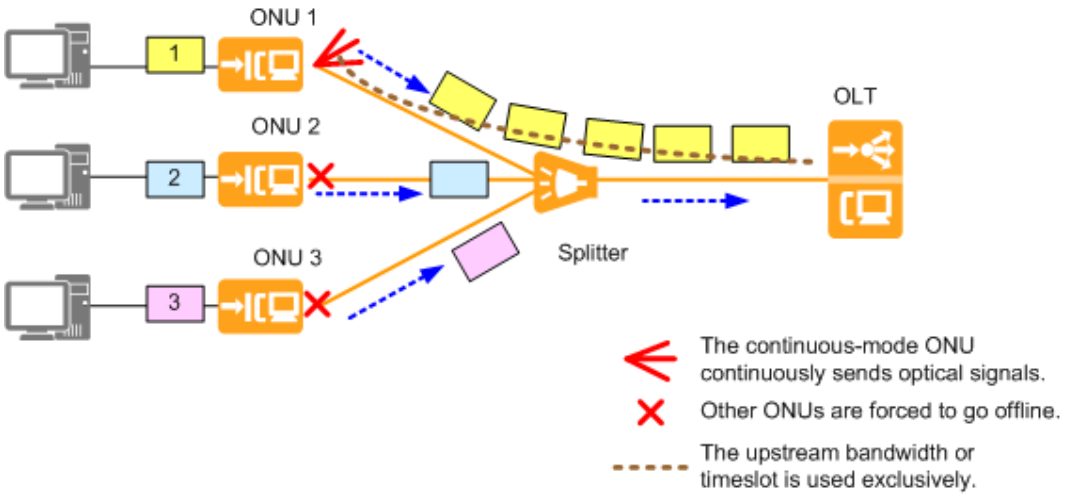
PON networks use the P2MP(point-to-multipoint) network architecture. They use time division multiple access (TDMA) in the upstream direction. ONUs must send optical signals upstream at the timeslots allocated by the OLT to prevent data conflict.

A continuous-mode ONU is also called a continuous-mode rogue ONU. It continuously sends optical signals to the upstream direction.

A continuous-mode ONU adversely affects the system as follows:

- If this ONU has been online, some or all ONUs connected to the same PON port go offline or frequently go offline and online.
- If this ONU has not been configured, other ONUs that have not been configured and connected to the same PON port will not be discovered by the OLT.

Figure 2-35 Rogue ONU



Continuous-Mode ONU Detection

Continuous-mode ONU detection, also called rogue ONU detection, is used for detecting continuous-mode ONUs in the system and isolating them, ensuring proper system running. A continuous-mode ONU detection process involves three stages, checking, detection, and isolation.

Figure 2-36 Continuous-Mode ONU detection process

- The three stages are as follows: Checking: The OLT periodically checks whether a continuous-mode ONU connects to a PON port. This checking cannot locate the continuous-mode ONU.

 Copyright © Huawei Technologies Co., Ltd.

The OLT opens an empty window in the upstream direction of the PON to make all online ONUs stop sending upstream optical signals within a period of time, and detects ONU upstream optical signals. If the OLT receives optical signals, it then goes to the detection stage to locate the ONU.

- Detection: The OLT locates the continuous-mode ONU.

  The OLT issues a command to ONUs to instruct the optical modules of the ONUs to send optical signals upstream and checks whether optical signals can be received in the upstream direction. If other ONUs go offline after an ONU sends optical signals, this ONU is a continuous-mode ONU. In a detection process, the OLT checks all ONUs connected to a PON port for detecting all continuous-mode ONUs.
- Isolation: The OLT issues a command to power off the continuous-mode ONU, preventing this ONU from adversely affecting other ONUs connected to the same PON port.

  After an ONU is powered off by the OLT, the ONU cannot send optical signals upstream even after being reset or power recycled. This ONU can send optical signals upstream only after the OLT cancels the isolation.

NOTE

The OLT checks continuous-mode ONUs but does not detect or isolate them by default.

Handling a Continuous-Mode ONU

1. If an ONT goes online and other ONTs connected to the same PON port go offline or go online and offline frequently, or the **0x2e314021 There are illegal incursionary rogue ONTs under the port** alarm is reported to the OLT, a rogue ONT may exist in the system. In this case, locate the rogue ONT according to the following steps.

   NOTE

   You can also run the **display port state** command to query whether a rogue ONT exists under a PON port.
2. Run the **anti-rogueont manual-detect** command to detect, locate, and isolate a continuous-mode rogue ONT manually. Check whether the system generates alarm **0x2e314022 The ONT is rogue ONT** or **0x2e314021 There are illegal incursionary rogue ONTs under the port**.

   NOTE

   During rogue ONU detection, if the port of the ONU to be detected is configured with a type B protection group, you need to run the **force-switch** command to forcibly switch the protection group and then check for a rogue ONU to avoid protection switching. You can forcibly switch services to the work side for rogue ONU detection if you are not sure which feeder fiber functions properly. If the rogue ONU is not located on the working side, forcibly switch to the protection side to locate the rogue ONU. Then, run the **undo force-switch** command to cancel forced protection group switching.

   - If the **The ONT is rogue ONT** or **There are illegal incursionary rogue ONTs under the port** alarm is generated, a continuous-mode rogue ONT may exist. In this case, go to **3**.
   - If the **The ONT is rogue ONT** or **There are illegal incursionary rogue ONTs under the port** alarm is not generated, an irregular-mode rogue ONT may exist. In this case, go to **4**.

 Copyright © Huawei Technologies Co., Ltd.

3. Handle the ONT according to the generated alarm.
   - If the **0x2e314022 The ONT is rogue ONT** alarm is generated, replace the ONU. Then, go to **7**.
   - If the **0x2e314021 There are illegal incursionary rogue ONTs under the port** alarm is generated, go to **4**.

   NOTE

   If the **0x2e314021 There are illegal incursionary rogue ONTs under the port** alarm is generated, a continuous-mode rogue ONU may exist and this rogue ONU does not support Huawei-defined extended PLOAM messages, or the optical signal transmission of the ONU optical module cannot be controlled.
4. Run the **ont reset** command or the **ont deactivate** command to reset or deactivate ONUs under the PON port one by one. Then, check whether other ONUs that encounter the fault (going offline or going online and offline repeatedly) can go online.
   - If yes, it indicates that the ONU is a rogue ONU. Replace the ONU. Go to **7**.
   - If no, the ONT optical module may be damaged so that the rogue ONT fails to be reset or deactivated by running the command. In this case, go to **5**.
5. Locate a rogue ONU manually: Remove the upstream optical fibers of ONUs one by one from the optical splitter, and check whether other faulty ONUs go online normally.
   - If yes, it indicates that the ONT is a rogue ONT. Replace the ONU. Then, go to **7**.
   - If no, go to **6**.
6. Obtain technical support from Huawei.
7. The fault is rectified.

Limitations and Restrictions

- The OLT checks and analyzes the abnormality in the sending of upstream optical signal over a PON line, and identifies and isolates rogue ONUs of only non-malicious users. This feature does not apply to the intentionally sabotaged ONU or sub-standard ONU.
- A continuous-mode ONU (rogue ONU) is required to parse and respond to downstream PLOAM messages.
- The OLT can quickly locate a continuous-mode ONU only when the ONU supports standard PLOAM messages and the ONU optical module is correctly controlled. When an unconfigured ONU connected to a PON port is in the continuous mode, all the other unconfigured ONUs connected to this PON port will fail to be automatically discovered.

2.7 Remote Software Commissioning (PON)

2.7.1 Introduction

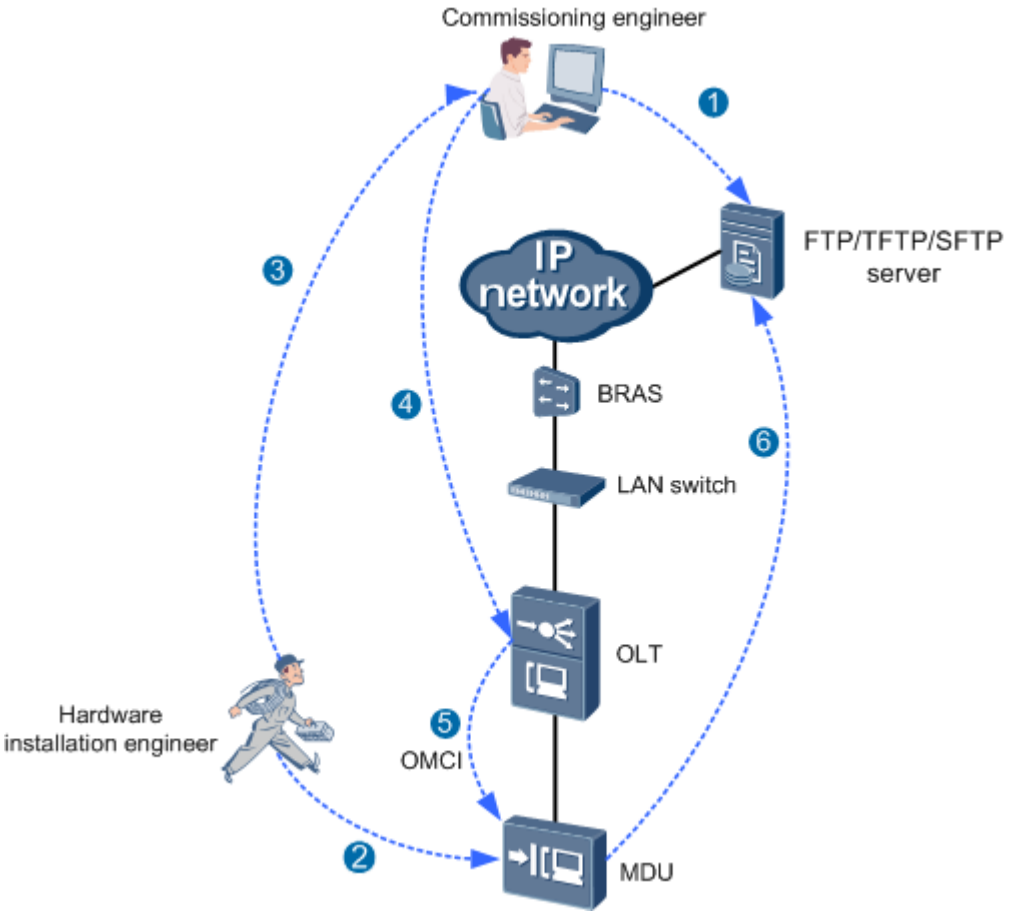
 Copyright © Huawei Technologies Co., Ltd.

During site deployment for a multi-dwelling unit (MDU) using PON upstream transmission, the MDU can be functional only after it is installed and manually commissioned by commissioning engineers onsite. To remove the need for onsite MDU commissioning, the MDU supports remote software commissioning. After the MDU is powered on, it automatically registers with the optical line terminal (OLT) and configures device data. This reduces site deployment costs.

2.7.2 Principles

Figure 2-37 shows the principles of remote software commissioning using PON upstream transmission.

Figure 2-37 Principles of remote software commissioning using PON upstream transmission



The process is as follows:

1. The commissioning engineer develops the automatic deployment policy file and configuration file, and uploads them to the FTP, TFTP or SFTP server.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

The automatic deployment policy file must comply with the xxx.xml naming format. The file must contain the device type, control board, protocol for transferring the configuration file, IP address of the server, and configuration file name.

The configuration file name must be of string type.

One automatic deployment policy file applies to all MDUs in one site. An example automatic deployment policy file used in one site is as follows:

```
<config>
<deploy-version value="1"/>
<product name="MA5xxx">                    //Device type
<mainboard name="H8xxxxxxx"/>              //Control board
<mainboard name="H8xxxxxxx"/>
<mainboard name="H8xxxxxxx"/>
<load>
<transfer-protocol isSupport="1" protocol="ftp" username="user" password="user123" port=""
serveripaddr="10.10.10.10"/>
<init-load-script>
<common-script isSupport="1" value="script.txt"/>
</init-load-script>
</load>
</product>
</config>
```

- **protocol**: indicates the protocol for transferring the configuration file, which can be FTP, SFTP, or TFTP.

  NOTE

  TFTP and FTP have security risks due to protocol limitations. You are advised to use SFTP.

- **username** and **password**: indicate the user name and password, respectively, when the configuration file is transferred using FTP or SFTP.
- **port**: specifies a port. This parameter is required to configure only if the default port used by the transfer protocol must be changed.
- **serveripaddr**: indicates the server IP address.
- **value**: specifies a configuration file name. When the configuration file is transferred using FTP or SFTP, the configuration file name may contain the path where this file is stored.

2. The hardware installation engineer obtains the MDU from the warehouse and delivers it to the site. Then, the hardware installation engineer installs the MDU hardware, connects lines for the MDU, and powers on the MDU.
3. The hardware installation engineer records and reports the SN of the MDU and site information to the commissioning engineer.
4. A commissioning engineer adds an ONT offline, and configures the ONT IP address, service flows, and automatic deployment profile on the OLT.
5. The OLT uses optical network terminal management and control interface (OMCI) to send the path where the automatic deployment policy file is stored to the MDU.
6. After being powered on, the MDU receives the path where the automatic deployment policy file is stored and starts automatic device deployment.

   NOTE

   The automatic device deployment takes effect on the MDU only if the MDU starts from an empty database. If the MDU database is not empty, run the **erase flash data** command to clear the database, or run the **load data** command to load an empty database to the MDU.

 Copyright © Huawei Technologies Co., Ltd.

7. The MDU requests for the automatic deployment policy file from the FTP, TFTP or SFTP server and implements automatic device configuration based on the automatic deployment policy specified in the file.

2.7.3 Configuring Remote Software Commissioning (GPON)

The MDU supports remote software commissioning using GPON upstream transmission. After the MDU is powered on, it automatically registers with the OLT and configures device data.

Procedure

Step 1 Run the **ont add** command add an ONT in offline mode.

Step 2 Run the **rn ipconfig** command to set the IP address of this MDU.

Step 3 Run the **service-port** command to create service flows.

Step 4 Run the **rn deploy-profile add** command to configure an automatic deployment policy profile.

NOTE

In remote software commissioning, the **terminal user authentication-mode AAA** *domain-name* command needs to be set at the last of the configuration file. Otherwise, this command configuration fails to be issued.

Step 5 Run the **rn deploy-config** command to bind the configured profile to the remote network (RN) equipment.

Step 6 Run the **display rn deploy log** command to query automatic deployment results and failure causes if the deployment fails.

----End

Example

Example:

1. On port 0/14/0, set the IP address of the RN with ONT ID 0 to 192.168.1.33, subnet mask to 255.255.255.0, gateway IP address to 192.168.1.1, management VLAN ID to 1, and priority to 3.
2. Add an automatic deployment policy profile and bind the profile to port 0/14/0, where the profile ID is 1, file name is deploy-backup.xml, file server IP address is 10.10.10.10, file transfer protocol is SFTP, user name is user, and password is user123.

```
huawei(config)#interface gpon 0/14
huawei(config-if-gpon-0/14)#port 0 ont-auto-find enable
huawei(config-if-gpon-0/14)#ont add 0 sn-auth 485754437B6F5130 snmp ont-lineprofile-id 1
huawei(config-if-gpon-0/14)#quit
huawei(config)#rn ipconfig 0/14/0 0 ip-address 192.168.1.33 mask 255.255.255.0 gateway 192.168.1.1 vlan 1 priority 3
huawei(config)#service-port vlan 1 gpon 0/14/0 ont 0 gemport 0 multi-service user-vlan 1
huawei(config)#rn deploy-profile add profile-id 1 filename deploy-backup.xml ip 10.10.10.10 sftp user
huawei(config)#rn deploy-config 0/14/0 0 profile-id 1
```

 Copyright © Huawei Technologies Co., Ltd.

2.8 GPON Configuration Guide (Profile mode)

GPON configurations include the configurations on GPON profiles, ONTs, and ports. The following section describes configuration methods.

Context

The xPON mode includes two types: distributed (discrete) mode and profile mode. The differences between the two modes are as follows:

- In the distributing mode, ONTs cannot be added in batches. Instead, ONTs need to be configured one by one.
- In profile mode, you can pre-configure ONT line profiles and ONT service profiles and bind ONTs of the same configurations to the same profile to add them in batches, which significantly improves service provisioning efficiency.

The xPON mode is determined during site provisioning and will not be changed.

You can run the **display xpon mode** command to query the xPON mode of the current system.

2.8.1 Configuring a GPON ONT Profile (Profile mode)

In distributed mode, GPON ONT profiles include the GPON ONT capability profile and the GPON ONT alarm profile. In profile mode, GPON ONT profiles include DBA profiles, line profiles, service profiles, and alarm profiles. This topic describes how to configure these profiles.

Context

GPON ONT profiles contain the parameters required for configuring the GPON access service, of which,

- DBA profiles specify GPON traffic parameters. The DBA profile bound to an OLT enables the OLT to dynamically allocate bandwidths, improving upstream bandwidth utilization.
- In profile mode, the line profile is mainly used to configure the information related to DBA, T-CONT, and GEM port. The service profile is used to configure the actual ONT capability and the parameters related to services. The line profile is mandatory and the service profile is optional and dependent of service requirements. Set related attributes in line profile mode and service profile mode, and directly bind the ONT to the line profile and service profile.
- The GPON ONT alarm profile provides a series of alarm threshold parameters that are used for performance measurement and monitoring of activated ONU lines. After a GPON alarm profile is bound to an ONU, the ONU sends alarms to the log host and the NMS if the performance statistics of the line exceed the threshold that is specified in the profile.

NOTE

In this document, ONUs include MDUs and ONTs.

 Copyright © Huawei Technologies Co., Ltd.

2.8.1.1 Configuring a DBA Profile

A DBA profile defines the traffic parameters of xPON and can be bound to a T-CONT to dynamically allocate the bandwidth and improve the usage of the upstream bandwidth.

Default Configuration

Table 2-6 lists the default settings of the DBA profiles.

Table 2-6 Default settings of the DBA profiles

Parameter	Default Setting	Remarks
Default DBA profile ID in the system	0-9	You can run the **display dba-profile all** command to query the parameter values of each default DBA profile.

Procedure

Step 1 Add a DBA profile.

Run the **dba-profile add** command to add a DBA profile. By default, the system provides DBA profiles 0–9 with typical traffic parameter values. The default profiles cannot be added or deleted.

NOTE

- By default, a T-CONT is not bound to any DBA profile. You must bind a T-CONT to a DBA profile.
- When you add a DBA profile, the bandwidth value must be a multiple of 64. If you enter a bandwidth value not of a multiple of 64, the system adopts the closest multiple of 64 that is smaller than the value you enter.

Step 2 Query a DBA profile.

Run the **display dba-profile** command to query a DBA profile.

----End

Example

To add a DBA profile with the type of type3, set the profile name to DBA_100M and the required bandwidth to 100 Mbit/s, do as follows:

```
huawei(config)#dba-profile add profile-name DBA_100M type3 assure 102400 max 102400
huawei(config)#display dba-profile profile-name DBA_100M
```

2.8.1.2 Configuring a GPON ONT Line Profile (Profile Mode)

This topic describes how to configure a GPON ONT line profile and use it when adding an ONT. When an ONT is managed by OMCI or SNMP, the ONT must be bound to a GPON ONT line profile.

 Copyright © Huawei Technologies Co., Ltd.

Default Configuration

Table 2-7 lists the default settings of a GPON ONT line profile.

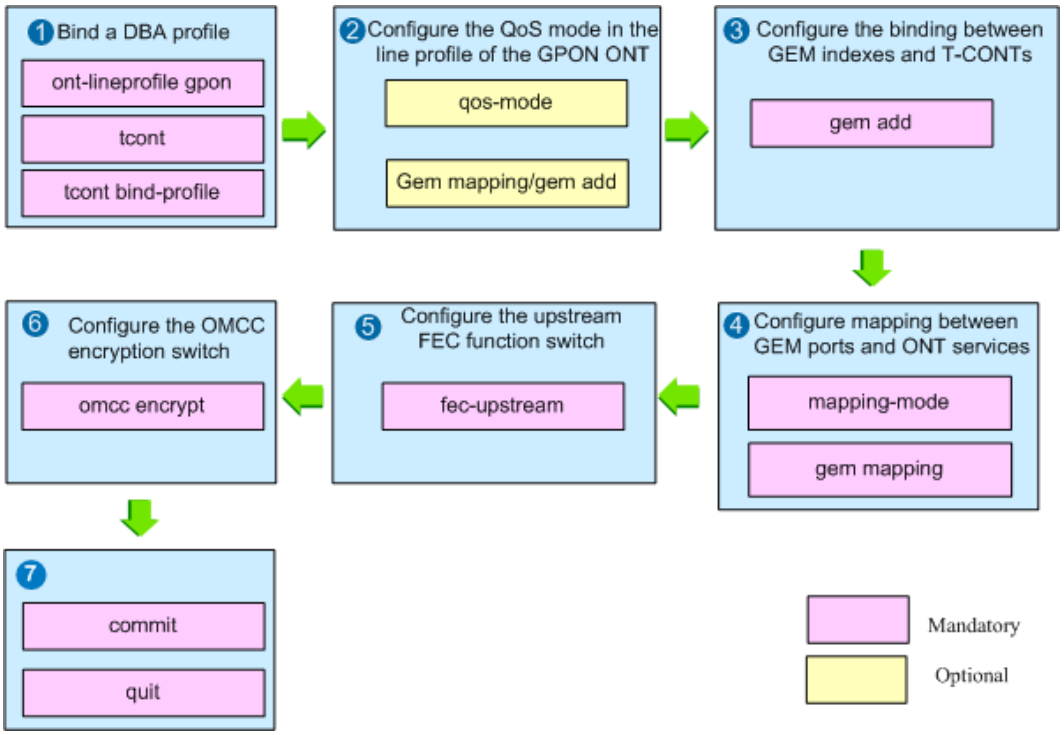
Table 2-7 Default settings of a GPON ONT line profile

Parameter	Default Setting
QoS mode	Priority-queue (PQ) scheduling mode
Mapping mode supported by the ONT	VLAN mapping mode
Upstream FEC switch	Disabled

Configuration Process

Figure 2-38 shows the process of configuring a GPON ONT line profile.

Figure 2-38 Process of configuring a GPON ONT line profile



Procedure

Step 1 Bind a T-CONT to a DBA profile.

1. Run the **ont-lineprofile gpon** command to add a GPON ONT line profile and enter the GPON ONT line profile mode.

   Regardless of whether the ONT is in the OMCI or SNMP management mode, the line profile must be configured for the ONT. After adding a GPON ONT

 Copyright © Huawei Technologies Co., Ltd.

line profile, directly enter the GPON ONT line profile mode to configure the related attributes of the ONT line.

2. Use the following two methods to bind a DBA profile. Select either method as required. Both methods can coexist in the system.
   - In line profile mode:

     This method is applicable to the scenario where the DBA profile is stable and the terminals are of a single type.

     Run the **tcont** command to bind the T-CONT to a DBA profile. Ensure that **2.8.1.1 Configuring a DBA Profile** is completed before the configuration.
   - In GPON mode:

     This method is applicable to the scenario where the DBA profile changes frequently and the terminals are of different types.

     i. Run the **tcont** command to create a T-CONT, which is not bound to the DBA.

     ii. After the configuration of a GPON ONT line profile is complete, enter the GPON mode. Run the **tcont bind-profile** command to bind the T-CONT to a DBA profile. Ensure that **2.8.1.1 Configuring a DBA Profile** is completed before the configuration.

   NOTE

   By default, T-CONT 0 of an ONT is used by OMCI and is bound to DBA profile 1. The configuration suggestions for the OMCI T-CONT are as follows:
   - Do not modify the DBA profile bound to T-CONT 0. If you need to modify the profile, ensure that the fixed bandwidth of the modified profile is not lower than 1 Mbit/s.
   - Do not bind a GEM port to the T-CONT. That is, ensure that the T-CONT does not carry any service.
   - If the sum of the fixed bandwidth and assured bandwidth of the bound DBA profile is larger than the remaining bandwidth of the GPON port, the binding fails and the system displays a message "Failure: The bandwidth is not enough". In this case, you can run the **display port info** command to query the remaining bandwidth (Left guaranteed bandwidth (kbit/s)) of the GPON port, and then decrease the fixed bandwidth and assured bandwidth of the bound DBA profile accordingly.

Step 2 (Optional) Configure the QoS mode of the GPON ONT line profile.

Run the **qos-mode** command to configure the QoS mode of the GPON ONT line profile to be the same as the QoS mode of the GEM port. By default, the QoS mode of the ONT line profile is the PQ scheduling mode. The three QoS modes are as follows:

- flow-car: When this mode is selected, **flow-car** should be selected in the **gem mapping** command, and the maximum traffic depends on the traffic profile bound to the service port. Run the **traffic table ip** command to create a required traffic profile before the configuration.
- gem-car: When this mode is selected, **gem-car** should be selected in the **gem add** command, and the maximum traffic depends on the traffic profile bound to the GEM port.
- priority-queue: When this mode is selected, **priority-queue** should be selected in the **gem add** command. The system has eight default queues (0-7). Queue 7 has the highest priority and the traffic of this queue must be ensured first.

 Copyright © Huawei Technologies Co., Ltd.

The maximum traffic depends on the DBA profile bound to the corresponding T-CONT.

Step 3 Configure the binding relationship between the GEM index and the T-CONT.

Run the **gem add** command to configure the binding relation between the GEM index and the T-CONT in the GPON ONT line profile.

The ONT can carry services only after the mapping between the GEM port and the T-CONT, and the mapping between the GEM port and the service port are configured for the ONT. A correct attribute should be selected for **service-type** based on the service type. Select **eth** when the Ethernet service is carried. Select **tdm** when the TDM service is carried.

Step 4 Configure the mapping between the GEM port and the ONT-side service.

Run the **gem mapping** command to set up the mapping between the GEM port and the ONT-side service.

Before the configuration, run the **mapping-mode** command to configure the mapping mode supported by the ONT to be the same as the configured mapping mode between the GEM port and the ONT-side service. By default, the ONT supports the VLAN mapping mode.

Mapping mode description

Parameter	Description	
	Mapping Mode	**Mandatory Configuration Item for the gem mapping Command**
vlan	Indicates the VLAN mapping. The VLAN mapping is the default mapping mode.	**vlan**
priority	Indicates the 802.1p priority mapping.	**priority**
vlan-priority	Indicates the VLAN + 802.1p priority mapping.	**vlan** and **priority**
port	Indicates the port ID mapping.	*ont-portid*
port-vlan	Indicates the port ID+VLAN mapping.	*ont-portid* and **vlan**
port-priority	Indicates the port ID+802.1p priority mapping.	*ont-portid* and **priority**
port-vlan-priority	Indicates the port ID+VLAN +802.1p priority mapping.	*ont-portid*, **vlan**, and **priority**

 Copyright © Huawei Technologies Co., Ltd.

Parameter	Description	
	Mapping Mode	**Mandatory Configuration Item for the gem mapping Command**
iptos	Indicates the IP ToS field mapping. IP ToS is the ToS priority of the IP packet header. This mapping is based on the priority of the ToS domain and is considered as the extension of the 802.1p priority.	**priority**
vlan-iptos	Indicates the VLAN+IP ToS field mapping.	**vlan** and **priority**

- The mapping modes of the ETH port and the MOCA port are as follows:
  - If the port is specified and then the VLAN is further specified, the mapping mode should be configured to **port-vlan** in the **mapping-mode** command. That is, the port+VLAN mapping mode is used.
  - If the port is specified and then the priority is further specified, the mapping mode should be configured to **port-priority** in the **mapping-mode** command. That is, the port+priority mapping mode is used.
  - If the port and the VLAN are specified and then the priority is further specified, the mapping mode should be configured to **port-vlan-priority** in the **mapping-mode** command. That is, the port+VLAN+priority mapping mode is used.
- As a special port, the IPHOST or E1 port is not restricted by the ONT mapping mode.

When the mapping mode is **vlan-priority** or **port-vlan-priority**,

- If a GEM port is mapped to multiple VLANs, any of these VLANs cannot map to any other GEM port.
- If a VLAN is mapped to multiple GEM ports, any of these GEM ports cannot map to any other VLAN.

Step 5 Configure the upstream FEC switch.

Run the **fec-upstream** command to configure the upstream FEC switch of the GPON ONT line profile. By default, this switch is disabled.

In the FEC check, the system inserts redundancy data into normal packets. In this way, the line has certain error tolerant function, but certain bandwidth resources are wasted. Enabling the FEC function enhances the error tolerant capability of the line but occupies certain bandwidth. Therefore, determine whether to enable the FEC function based on the actual line planning.

Step 6 Configure the status of the ONT Management and Control Channel (OMCC) encryption switch.

 Copyright © Huawei Technologies Co., Ltd.

Run the **omcc encrypt** command to configure the status of the OMCC encryption switch. When OMCC encryption is enabled, OMCC packets are encrypted; otherwise, OMCC packets are not encrypted. By default, this switch is disabled.

Step 7 Run the **commit** command to make the parameters of the profile take effect. The configuration of a line profile takes effect only after you perform this operation.

NOTE

If this profile is not bound, all the parameters that are configured take effect when the profile is bound. If this profile is already bound, the configuration takes effect on all ONTs bound to this profile immediately.

Step 8 Run the **quit** command to return to the global configuration mode.

----End

Example

Example: Add GPON ONT line profile 5. Create a channel for carrying the ETH service. Set T-CONT to 1, bind DBA profile 12, set GEM index to 1, bind it to T-CONT 1, and map it to ETH port 1 of the ONT. Use the QoS policy of controlling the traffic based on GEM ports, and bind the GEM port to default traffic profile 6.

```
huawei(config)#ont-lineprofile gpon profile-id 5
huawei(config-gpon-lineprofile-5)#tcont 1 dba-profile-id 12
huawei(config-gpon-lineprofile-5)#qos-mode gem-car
huawei(config-gpon-lineprofile-5)#gem add 1 eth tcont 1 gem-car 6
huawei(config-gpon-lineprofile-5)#mapping-mode port
huawei(config-gpon-lineprofile-5)#gem mapping 1 0 eth 1
huawei(config-gpon-lineprofile-5)#commit
huawei(config-gpon-lineprofile-5)#quit
```

To modify GPON ONT line profile 5, and change the DBA profile bound to T-CONT 1 from DBA profile 12 to DBA profile 10, do as follows:

```
huawei(config)#ont-lineprofile gpon profile-id 5
huawei(config-gpon-lineprofile-5)#tcont 1 dba-profile-id 10
huawei(config-gpon-lineprofile-5)#commit
huawei(config-gpon-lineprofile-5)#quit
```

To modify GPON ONT line profile 5, bind GEM index 1 to T-CONT 2 (originally bound to T-CONT 3), and map GEM index 1 to ETH port 2 on the ONT, do as follows:

NOTE

- To modify the binding between a GEM index and a T-CONT, you need to delete the existing binding entry and then bind the GEM index to the new T-CONT.
- If the GEM index is used by a service port, run the **undo service-port** command to delete the service port before deleting the GEM index.

```
huawei(config)#ont-lineprofile gpon profile-id 5
huawei(config-gpon-lineprofile-5)#undo gem mapping 1 0
huawei(config-gpon-lineprofile-5)#gem delete 1
huawei(config-gpon-lineprofile-5)#gem add 1 eth tcont 2
huawei(config-gpon-lineprofile-5)#gem mapping 1 0 eth 2
huawei(config-gpon-lineprofile-5)#commit
huawei(config-gpon-lineprofile-5)#quit
```

2.8.1.3 Configuring a GPON ONT Service Profile (Profile Mode)

The GPON ONT service profile provides a channel for configuring the service of the ONT managed in the OMCI mode. The ONT (such as the MDU) managed in

 Copyright © Huawei Technologies Co., Ltd.

the SNMP mode does not support the configuration of the GPON ONT service profile. To configure the service of the ONT (such as the MDU) managed in the SNMP mode, you need to log in to the ONT.

Default Configuration

Table 2-8 lists the default settings of the GPON ONT service profile.

Table 2-8 Default settings of the GPON ONT service profile

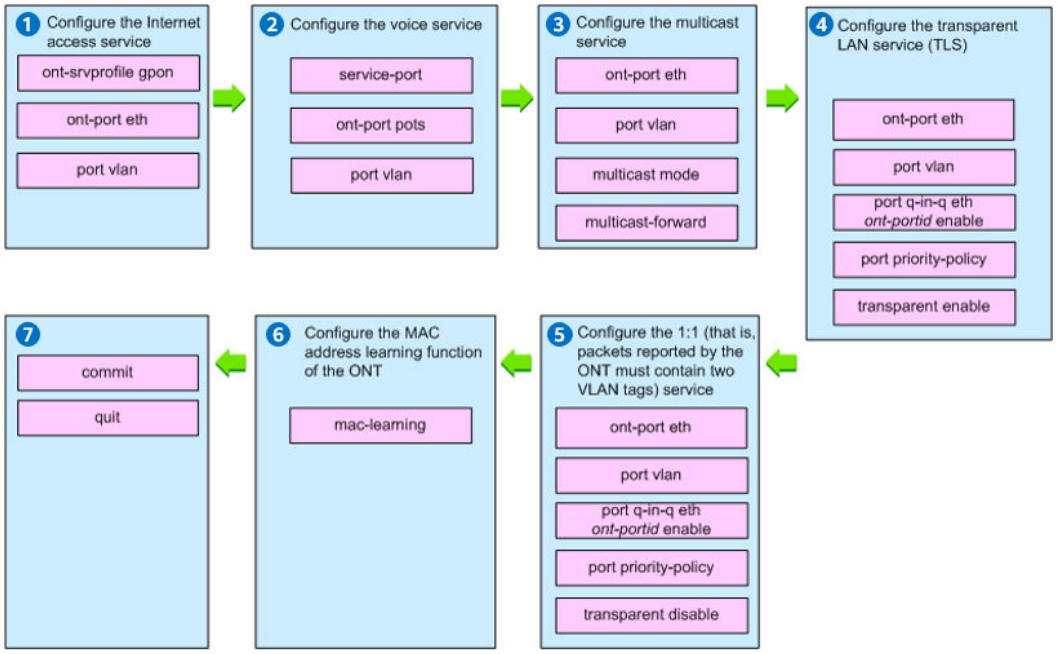
Parameter	Default Setting
Multicast mode of the ONT	Unconcern (the OLT does not perform any processing)
Mode for the ONT to process the VLAN tag of the multicast data packets	Unconcern
Source of the priority copied for the upstream packets on the ONT port	Unconcern
QinQ attribute for the Ethernet port of the ONT	Unconcern
Transparent transmission function of the ONT	Disabled
MAC address learning function of the ONT	Enabled

Configuration Process

Figure 2-39 shows the process of configuring a GPON ONT service profile.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-39 Process of configuring a GPON ONT service profile



Procedure

Step 1 Configure the Internet access service.

1. Run the **ont-srvprofile gpon** command to add a GPON ONT service profile, and then enter the GPON ONT service profile mode. If the ONT management mode is the SNMP mode, you do not need to configure the service profile. After adding a GPON ONT service profile, directly enter the GPON ONT service profile mode to configure the related items. Select the configuration items according to the service requirements.
2. Run the **ont-port eth** command to configure the port capability set of the ONT. The capability set plans the number of various ports supported by the ONT. The port capability set of the ONT must be the same as the actual capability set of the ONT.

   If the port capability set of an ONT is set to **adaptive**, the system automatically adapts to the actual capabilities of the online ONT. By default, 8 ETH ports and 1 IPHOST port are displayed.
3. Run the **port vlan(gpon profile)** command to configure the port VLAN of the ONT.

   VLAN Configuration Description

 Copyright © Huawei Technologies Co., Ltd.

Table 2-9 Packet Processing Policy

Transparent transmission based on the port	Transparently transmit the tagged packet directly.	Add a native VLAN tag to the untagged packet.	Transparently transmit the untagged packet directly.
QinQ based on the port	Add a native VLAN tag to the tagged packet. **NOTE** The native VLAN item must be set to concern by running the **native vlan(gpon)** command. You can run the **display ont-srvprofile gpon** command to query the status of the native VLAN item.	Add a native VLAN tag to the untagged packet.	Add a native VLAN tag to the untagged packet.
Translation based on the VLAN	– When only parameter *SVLAN* is configured, it indicates that CVLAN is the same as SVLAN. – If the VLAN of the packet is *user-vlanid*, the VLAN of the packet will be translated. – Otherwise, the packet will be discarded.	The VLAN translation pair for untagged packets cannot be configured when the native VLAN is enabled. A native VLAN tag is added to the untagged packet and the VLAN translation pair will not be checked. That is, regardless of whether the native VLAN is in the VLAN translation pair or not, packets are allowed to pass through the VLAN.	– The native VLAN cannot be enabled for the whole ONT when the VLAN translation pair for untagged packets is configured. – Otherwise, the packet will be discarded.

– The transparent transmission mode of a port conflicts with the transparent transmission mode of the VLAN of the port. That is, after

 Copyright © Huawei Technologies Co., Ltd.

setting the transparent transmission mode of a port, you need not set the transparent transmission mode of the VLAN of the port.

- In single-bridge mode, the transparent transmission mode of the VLAN of multiple ports must be set to be the same as the transparent transmission mode of these ports. In multiple-bridge mode, the transparent transmission modes of the VLAN and the ports can be different.
- When you configure the transparent transmission mode of the untagged packets by using QinQ, this command cannot be directly used to configure the transparent transmission mode of the untagged packets.
- The Ethernet port and VDSL port of an ONT can be configured to add a VLAN QinQ tag to untagged packets on the user side based on the packet encapsulation type. In this case, the Native VLAN item in the **ont native-vlan(distributing-mode)** command must be set to unconcern.
- Configure whether the ONT concerns the Native VLAN in the ONT service profile. By default, the ONT concerns the Native VLAN item.
  - You can run the **ont port native-vlan(gpon)** command to configure the native VLAN of an ONT port only when the native VLAN is set to concern. If the native VLAN of an ONT port is set to a non-default value, the native VLAN cannot be set to unconcern. To query the VLAN configuration of an ONT port, run the **display ont port vlan(distributing-mode)** command.
  - You can run the **port q-in-q** command to configure the QinQ attribute of the Ethernet port on an ONT only when the Native VLAN is set to concern. You cannot change the Native VLAN item to unconcern when the QinQ attribute of the Ethernet port on the ONT is enabled. You can run the **display ont-srvprofile gpon** command to query the QinQ attribute of an ONT port.
  - You can run the **ont port vlan(distributing-mode)** command to configure the policy of processing untagged packets on an ONT port only when the Native VLAN is set to unconcern. When the policy of processing untagged packets on an ONT port is configured, the Native VLAN item cannot be set to concern. You can run the **display ont port vlan(distributing-mode)** command to query the policy for processing untagged packets on an ONT port.

Step 2 Configure the voice service.

NOTE

The voice service of the ONT is configured by issuing an XML file to the NMS and the OLT performs only transparent transmission. You only need to run the **service-port** command to create a service port carrying the voice service.

1. Run the **ont-port pots** command to configure the port capability set of the ONT. The port capability set in the ONT service profile must be the same as the actual ONT capability set.

   If the port capability set of an ONT is set to **adaptive**, the system automatically adapts to the actual capabilities of the online ONT. By default, 8 ETH ports and 1 IPHOST port are displayed.

 Copyright © Huawei Technologies Co., Ltd.

2. Run the **port vlan(gpon profile)** command to configure the port VLAN of the ONT.

   VLAN Configuration Description

Step 3 Configure the multicast service.

1. Run the **ont-port eth** command to configure the port capability set of the ONT. The port capability set in the ONT service profile must be the same as the actual ONT capability set.

   If the port capability set of an ONT is set to **adaptive**, the system automatically adapts to the actual capabilities of the online ONT. By default, 8 ETH ports and 1 IPHOST port are displayed.

2. Run the **port vlan(gpon profile)** command to configure the port VLAN of the ONT.

   VLAN Configuration Description

3. Run the **multicast mode** command to configure the multicast mode of the ONT. By default, the multicast mode of the ONT is **unconcern**.
   - Unconcern: indicates the unconcern mode. After this mode is selected, the OLT does not limit the multicast mode, and the multicast mode on the OLT automatically matches the multicast mode on the ONT.
   - Igmp-snooping: IGMP snooping obtains the related information and maintains the multicast forwarding entries by listening to the IGMP packets in the communication between the user and the multicast router.
   - Olt-control: indicates the dynamic controllable multicast mode. A multicast forwarding entry can be created for the multicast join packet of the user only after the packet passes the authentication. This mode is supported by the MDU, but is not supported by the ONT.
4. Run the **multicast-forward** command to configure the processing mode on the VLAN tag of the multicast data packets for the ONT. By default, the multicast forwarding mode of the ONT is **unconcern**.
   - Unconcern: indicates the unconcern forwarding mode. After this mode is selected, the OLT does not process the VLAN tag of the multicast data packets.
   - tag: Sets the multicast forwarding mode to with VLAN tag. To transparently transmit the VLAN tag of the multicast packets, select **transparent**. To translate the VLAN tag of multicast packets, select **translation** and configure the translated VLAN tag.
   - Untag: Set the multicast forwarding mode not to contain the VLAN tag.

Step 4 Configure the transparent LAN service (TLS).

1. Run the **ont-port eth** command to configure the port capability set of the ONT. The port capability set in the ONT service profile must be the same as the actual ONT capability set.

   If the port capability set of an ONT is set to **adaptive**, the system automatically adapts to the actual capabilities of the online ONT. By default, 8 ETH ports and 1 IPHOST port are displayed.

2. Run the **port vlan(gpon profile)** command to configure the port VLAN of the ONT.

 Copyright © Huawei Technologies Co., Ltd.

VLAN Configuration Description

3. Run the **port q-in-q eth ont-portid enable** command to enable the QinQ function of the Ethernet port on the ONT. By default, the QinQ function of the Ethernet port on the ONT is unconcerned.
4. Run the **port priority-policy** command to configure the source of the priority copied for the upstream packets on the ONT port. By default, the source of the priority copied for the upstream packets on the ONT Ethernet port is unconcerned.
   - unconcern: The upstream priority copy source of the ONT Ethernet port is not concerned.
   - assigned: Specifies the priority. Run the **ont port native-vlan** command to specify the **priority** of the port.
   - copy-cos: Copy the priority. Copy the priority from C-TAG.
5. Run the **transparent enable** command to enable the transparent transmission function of the ONT. By default, the transparent transmission function of the ONT is disabled. After the transparent transmission function of the ONT is enabled, all packets (including service packets and protocol packets) are transparently transmitted by the ONT.

NOTE

The service port for the TLS service must also be of the TLS type. Run the **service-port** command to create a service port of the TLS type. Select **other-all** for the multi-service type.

Step 5 Configure the 1:1 (that is, packets reported by the ONT must contain two VLAN tags) service.

1. Run the **ont-port eth** command to configure the port capability set of the ONT. The port capability set in the ONT service profile must be the same as the actual ONT capability set.

   If the port capability set of an ONT is set to **adaptive**, the system automatically adapts to the actual capabilities of the online ONT. By default, 8 ETH ports and 1 IPHOST port are displayed.
2. Run the **port vlan(gpon profile)** command to configure the port VLAN of the ONT.

VLAN Configuration Description

3. Run the **port q-in-q eth ont-portid enable** command to enable the QinQ function of the Ethernet port on the ONT. By default, the QinQ function of the Ethernet port on the ONT is unconcerned.
4. Run the **port priority-policy** command to configure the source of the priority copied for the upstream packets on the ONT port. By default, the source of the priority copied for the upstream packets on the ONT Ethernet port is unconcerned.
   - Unconcern: The source of the priority copied for the upstream packets on the Ethernet port of the ONT is not concerned.
   - assigned: Specifies the priority. Run the **ont port native-vlan** command to specify the **priority** of the port.
   - Copy-cos: Copy the priority. Copy the priority from C-TAG.
5. Run the **transparent disable** command to disable the transparent transmission function of the ONT.

 Copyright © Huawei Technologies Co., Ltd.

Step 6 Run the **mac-learning** command to configure the MAC address learning function of the ONT. This function is enabled by default.

Step 7 Run the **commit** command to make the parameters of the profile take effect. The configuration of a service profile takes effect only after you perform this operation.

NOTE

If this profile is not bound, all the parameters that are configured take effect when the profile is bound. If this profile is already bound, the configuration takes effect on all ONTs bound to this profile immediately.

Step 8 Run the **quit** command to return to the global config mode.

----End

Example

Add GPON ONT service profile 5 for the Internet access service. The ONT supports 4 ETH ports. The VLAN ID of the ETH port is 10. Do as follows:

```
huawei(config)#ont-srvprofile gpon profile-id 5
huawei(config-gpon-srvprofile-5)#ont-port eth adaptive
huawei(config-gpon-srvprofile-5)#port vlan eth 1-4 10
huawei(config-gpon-srvprofile-5)#commit
huawei(config-gpon-srvprofile-5)#quit
```

Add GPON ONT service profile 6 for the multicast service. The ONT supports 4 ETH ports. The VLAN ID of the ETH port is 100. The multicast mode of the ONT is the controllable multicast mode. The STB supports only VLAN tag 841, and multicast VLAN tag needs to be translated. Do as follows:

```
huawei(config)#ont-srvprofile gpon profile-id 6
huawei(config-gpon-srvprofile-6)#ont-port eth adaptive
huawei(config-gpon-srvprofile-6)#port vlan eth 1-4 100
huawei(config-gpon-srvprofile-6)#multicast mode olt-control
huawei(config-gpon-srvprofile-6)#multicast-forward tag translation 841
huawei(config-gpon-srvprofile-6)#commit
huawei(config-gpon-srvprofile-6)#quit
```

2.8.1.4 Configuring a GPON ONT Alarm Profile

This topic describes how to add an alarm profile, and configure most of the performance parameters for various ONT lines as a profile. After the alarm profile is configured and bound successfully, the ONT can directly use the profile when it is activated.

Context

An ONT alarm profile defines a series of alarm thresholds that are used to monitor the performance of an activated ONT line. When the statistics result of a parameter reaches the alarm threshold, the NE is notified and an alarm is sent to the log server and the NMS.

The system has a default alarm threshold profile with ID 1. This profile cannot be deleted but can be modified.

 Copyright © Huawei Technologies Co., Ltd.

Procedure

Step 1 Run the **gpon alarm-profile add** command to add a GPON ONT alarm profile.

If default alarm thresholds are used in a new alarm profile, the system does not report any alarm.

Step 2 Run the **display gpon alarm-profile** command to query the alarm profile.

----End

Example

To add GPON ONT alarm profile 5, set the alarm threshold for the packet loss of the GEM port to 10, set the alarm threshold for the number of mis-transmitted packets to 30, and use the default value 0 for all other thresholds, do as follows:

```
huawei(config)#gpon alarm-profile add profile-id 5
{ <cr>|profile-name<K> }:

  Command:
        gpon alarm-profile add profile-id 5
  Press 'Q' or 'q' to quit input
>  GEM port loss of packets threshold (0~100)[0]:10
>  GEM port misinserted packets threshold (0~100)[0]:30
>  GEM port impaired blocks threshold (0~100)0[0]:
>  Ethernet FCS errors threshold (0~100)[0]:
>  Ethernet excessive collision count threshold (0~100)[0]:
>  Ethernet late collision count threshold (0~100)[0]:
>  Too long Ethernet frames threshold (0~100)[0]:
>  Ethernet buffer (Rx) overflows threshold (0~100)[0]:
>  Ethernet buffer (Tx) overflows threshold (0~100)[0]:
>  Ethernet single collision frame count threshold (0~100)[0]:
>  Ethernet multiple collisions frame count threshold (0~100)[0]:
>  Ethernet SQE count threshold (0~100)[0]:
>  Ethernet deferred transmission count threshold (0~100)[0]:
>  Ethernet internal MAC Tx errors threshold (0~100)[0]:
>  Ethernet carrier sense errors threshold (0~100)[0]:
>  Ethernet alignment errors threshold (0~100)[0]:
>  Ethernet internal MAC Rx errors threshold (0~100)[0]:
>  PPPOE filtered frames threshold (0~100)[0]:
>  MAC bridge port discarded frames due to delay threshold (0~100)[0]:
>  MAC bridge port MTU exceeded discard frames threshold (0~100)[0]:
>  MAC bridge port received incorrect frames threshold (0~100)[0]:
>  CES general error time threshold(0~100)[0]:
>  CES severely time threshold(0~100)[0]:
>  CES bursty time threshold(0~100)[0]:
>  CES controlled slip threshold(0~100)[0]:
>  CES unavailable time threshold(0~100)[0]:
>  Drop events threshold(0~100)[0]:
>  Undersize packets threshold(0~100)[0]:
>  Fragments threshold(0~100)[0]:
>  Jabbers threshold(0~100)[0]:
>  Failed signal of ONT threshold(Format:1e-x, x: 3~8)[3]:
>  Degraded signal of ONT threshold(Format:1e-x, x: 4~9)[4]:
>  FEC uncorrectable code words threshold(0~1101600000)[0]:
>  FEC correctable code words threshold(0~1101600000)[0]:
>  Upstream PQ discarded byte alarm threshold(0~65535)[0]:
>  Downstream PQ discarded byte alarm threshold(0~65535)[0]:
>  Encryption key errors threshold(0~100)[0]:
>  XGEM key errors threshold(0~100)[0]:
>  XGEM HEC error count threshold(0~100)[0]:

  Adding an alarm profile succeeded
  Profile ID  : 5
  Profile name: alarm-profile_5
```

 Copyright © Huawei Technologies Co., Ltd.

2.8.2 Configuring a GPON ONT (Profile Mode)

The OLT provides end users with services through the ONT. The OLT can manage the ONT and the ONT can work in the normal state only after the channel between the OLT and the ONT is available.

Prerequisites

The GPON ONT profile is already created.

- For an ONT, **2.8.1.2 Configuring a GPON ONT Line Profile (Profile Mode)**, **2.8.1.3 Configuring a GPON ONT Service Profile (Profile Mode)**, and **2.8.1.4 Configuring a GPON ONT Alarm Profile** are already completed.
- For an MDU or ONU, **2.8.1.2 Configuring a GPON ONT Line Profile (Profile Mode)** and **2.8.1.4 Configuring a GPON ONT Alarm Profile** are already completed.

Context

The OLT uses the ONT Management and Control Interface (OMCI) protocol to manage and configure the GPON ONT, and supports the offline configuration of the ONT.

In the profile mode, the related configuration of the GPON ONT is already integrated in the service profile and the line profile. When adding an ONT, you only need to bind the ONT with the corresponding service profile and line profile.

Table 2-10 lists the default settings of the GPON ONT.

Table 2-10 Default settings of the GPON ONT

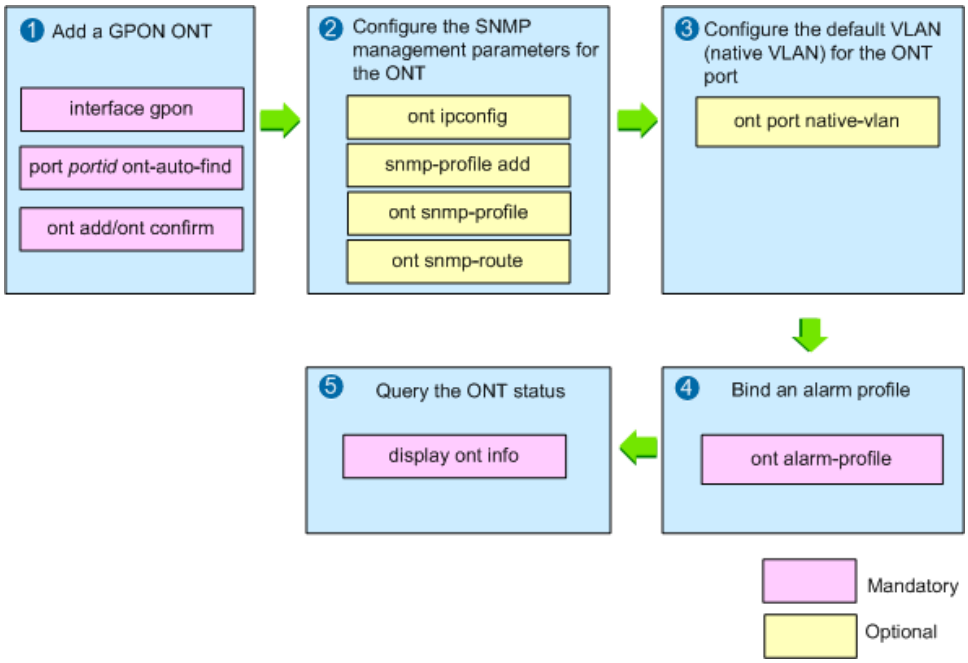
Parameter	Default Setting
ONT auto-find function of a GPON port	Disabled
ONT status after an ONT is added	Activated
Default VLAN of the ONT port	1

Configuration Process

Figure 2-40 shows the process of configuring a GPON ONT.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-40 Process of configuring a GPON ONT



Procedure

Step 1 Add a GPON ONT.

1. Run the **interface gpon** command to enter the GPON mode.
2. Run the **port portid ont-auto-find** command to enable the auto discovery function of the ONT. After the function is enabled, you can add an ONT according to the information reported by the system. By default, the ONT auto discovery function of a GPON port is disabled.

   NOTE

   An auto discovery ONT is in the auto discovery state. The auto discovery ONT can work in the normal state only after it is confirmed or added.
3. Run the **ont add** command to add an ONT offline, or run the **ont confirm** command to confirm the auto discovery ONT.

   When ONTs are added or confirmed, the system provides four authentication modes: SN, password, SN+password, LOID+CHECKCODE.

   - SN authentication: The OLT detects the serial number (SN) reported by an ONT. If the SN is consistent with the OLT configuration, authentication is passed and the ONT goes online. This mode requires recording all ONT SNs. Hence, it is used to confirm auto discovery ONTs and is not applicable to adding ONTs in batches.
   - Password authentication: The OLT detects the password reported by an ONT. If the password is consistent with the OLT configuration, the ONT goes online normally. This mode requires planning ONT passwords and does not require manually recording ONT SNs. Hence, it is applicable to adding ONTs in batches. The password authentication provides two discovery modes: always-on and once-on.

 Copyright © Huawei Technologies Co., Ltd.

    - always-on: After first password authentication is passed, no SN is allocated and password authentication is always used in subsequent authentications. This discovery mode is easy for future maintenance. In the always-on discovery mode, configurations do not need to be modified when an ONT is replaced and only the password is required. The always-on discovery mode has lower security. If other users know the password, the users will illegally have service permissions.
    - once-on: Password authentication is used for the first authentication. After the authentication is successful, the SN is automatically matched. The SN+password authentication mode is used for subsequent authentications. The ONT can go online only when both the password and SN are correct. The once-on authentication mode has high security. After an ONT is replaced or the password is mistakenly changed, the ONT needs to run the **ont re-discovery** command to be re-discovered, which requires more maintenance effort.
  - SN+password authentication: The OLT detects the password and SN reported by an ONT. If the password and SN are consistent with the OLT configuration, the ONT goes online normally. This authentication mode has the highest security but it requires manually recording ONT SNs.
  - LOID+CHECKCODE authentication: defined by a telecom operator. In this authentication mode, LOID has 24 bytes, and CHECKCODE has 12 bytes and is optional. Whether 24 bytes or 36 bytes are used for authentication depends on data planning, which is unified over the entire network. The OLT determines whether LOID+CHECKCODE reported by the ONT is the same as the configured one. If they are the same, the ONT authentication is passed. If they are different, the OLT obtains the ONT password and compares it with the last 10 bytes of the LOID. If they are the same, the ONT authentication is also passed. This operation is for compatibility with the ONTs using password authentication.

Adding ONTs in offline mode is applicable to the batch deployment scenario. All ONTs are added to the OLT to complete service provisioning beforehand. When a use subscribes to the service, an installation engineer takes an ONT to the user's house and completes configurations. After the ONT goes online and passes authentication, the service is provisioned. Generally, the password authentication mode or LOID authentication mode is used to confirm an ONT.

Adding ONTs in auto discovery mode is applicable to the scenario where a small number of ONTs are added. When users subscribe to the service, installation engineers take ONTs to the users' houses. After the ONTs go online, the OLT confirms the ONTs one by one. Generally, the MAC address authentication mode is used to confirm the ONTs.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

- If the ONU is an independent NE and is directly managed by the NMS through the SNMP management mode, select the SNMP management mode. For this mode, you only need to configure the parameters for the GPON line and the parameters for the management channel on the OLT. You only need to bind the ONU with a line profile.
- If the ONU is not an independent NE and all its configuration data is issued by the OLT through OMCI, select the OMCI management mode. For this mode, you need to configure all parameters (including line parameters, UNI port parameters, and service parameters) that are required for the ONU on the OLT. Configuring management channel parameters is not supported. You need to bind the ONT with a line profile and a service profile.
- Generally, the ONT management mode is set to the OMCI mode. You need to bind the ONT with a line profile and a service profile.

Step 2 (Optional) Configure the SNMP management parameters for the ONT. When the ONT management mode is the SNMP mode, you need to configure the SNMP management parameters for the ONT. The procedure is as follows:

1. Run the **ont ipconfig** command to configure the management IP address of the ONT. The IP address should not be in the same subnet for the IP address of the VLAN port.
2. Run the **ont snmp-profile** command to bind an SNMP profile to the ONT. Before running this command, ensure that an SNMP profile has been added by running the **snmp-profile add** command.
3. Run the **ont snmp-route** command to configure the static route of the NMS server. That is, configure the IP address of the next hop.

Step 3 Configure the default VLAN (native VLAN) for the ONT port.

Run the **ont port native-vlan** command to configure the default VLAN for the ONT port. By default, the default VLAN ID of the ONT port is 1.

- If the packets reported from a user (such a PC) to the ONT are untagged, the packets are tagged with the default VLAN of the port on the ONT and then reported to the OLT.
- If the packets reported from a user to the ONT are tagged, you need to configure the port VLAN of the ONT to be the same as the VLAN in the user tag. The packets are not tagged with the default VLAN of the port on the ONT but are reported to the OLT with the user tag.

Step 4 Bind an alarm profile.

Run the **ont alarm-profile** command to bind an alarm profile. Ensure that **2.8.1.4 Configuring a GPON ONT Alarm Profile** is completed before the configuration.

Step 5 Query the ONT status.

Run the **display ont info** command to query the ONT running status, configuration status, and matching status.

----End

Example

To add five ONTs in offline mode with password authentication mode (ONT passwords are 0100000001-0100000005), set the discovery mode of password

 Copyright © Huawei Technologies Co., Ltd.

authentication to always-on, and bind line profile 10 and service profile 10, do as follows:

```
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#ont add 0 password-auth 0100000001 always-on omci ont-lineprofile-id 10 ont-srvprofile-id 10
huawei(config-if-gpon-0/1)#ont add 1 password-auth 0100000002 always-on omci ont-lineprofile-id 10 ont-srvprofile-id 10
huawei(config-if-gpon-0/1)#ont add 2 password-auth 0100000003 always-on omci ont-lineprofile-id 10 ont-srvprofile-id 10
huawei(config-if-gpon-0/1)#ont add 3 password-auth 0100000004 always-on omci ont-lineprofile-id 10 ont-srvprofile-id 10
huawei(config-if-gpon-0/1)#ont add 4 password-auth 0100000005 always-on omci ont-lineprofile-id 10 ont-srvprofile-id 10
```

To add an ONT that is managed by the OLT through the OMCI protocol, confirm this ONT according to the SN 3230313185885B41 automatically reported by the system, and bind the ONT with line profile 3 and service profile 3 that match the ONT, do as follows:

```
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#port 0 ont-auto-find enable
huawei(config-if-gpon-0/1)#ont confirm 0 sn-auth 3230313185885B41 omci ont-lineprofile-id 3 ont-srvprofile-id 3
```

To add an ONU that is managed as an independent NE and whose SN is known as 3230313185885641, bind the ONU with line profile 4 that matches the ONU, configure the NMS parameters for the ONU, and set the management VLAN to 100, do as follows:

```
huawei(config)#snmp-profile add profile-id 1 v3 public private 10.10.10.53 161 bodyname
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#ont add 0 2 sn-auth 3230313185885641 snmp ont-lineprofile-id 4
huawei(config-if-gpon-0/1)#ont ipconfig 0 2 static ip-address 10.20.20.20 mask 255.255.255.0 gateway 10.10.20.1 vlan 100
huawei(config-if-gpon-0/1)#ont snmp-profile 0 2 profile-id 1
huawei(config-if-gpon-0/1)#ont snmp-route 0 2 ip-address 10.20.20.19 mask 255.255.255.0 next-hop 10.20.20.100
```

2.8.3 Configuring a GPON Port

To work normally and carry the service, a GPON port must be enabled first. This topic describes how to enable a GPON port and configure related attributes of the port.

Default Configuration

Table 2-11 lists the default settings of the GPON port.

Table 2-11 Default settings of the GPON port

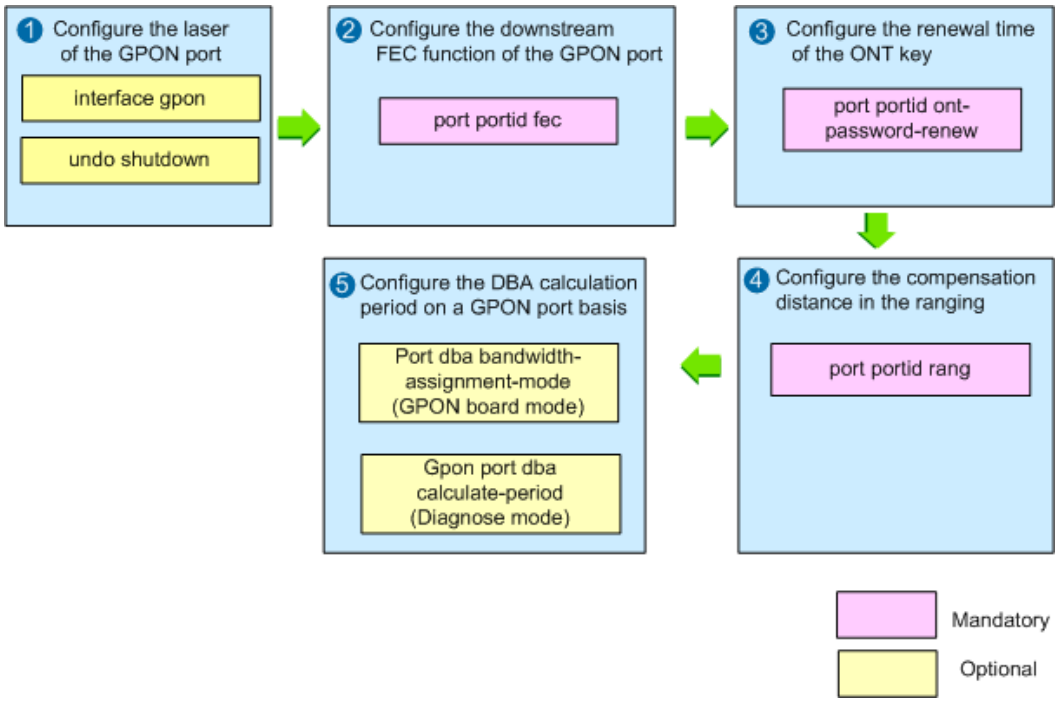
Parameter	Default Setting
GPON port	Enabled
Downstream FEC function of the GPON port	Disabled
Compensation distance range of the GPON port ranging	Minimum logical distance: 0 km; maximum logical distance: 20 km

 Copyright © Huawei Technologies Co., Ltd.

Configuration Process

Figure 2-41 shows the process of configuring a GPON Port.

Figure 2-41 Process of configuring a GPON Port



Procedure

Step 1 (Optional) Configure the optical port switch.

1. Run the **interface gpon** command to enter GPON mode.
2. Run the **undo shutdown** command to enable the laser of the GPON port. By default, the laser of the GPON port is enabled and the GPON port is available. In this case, skip this step.

NOTICE

- If an optical port is not required, run the **shutdown** command to disable the optical port.
- Disabling a PON port that carries services will cause the interruption of such services.

Step 2 Enables or disables the downstream FEC function of a port.

Run the **port portid fec** command to configure the FEC function of the GPON port. By default, the function is disabled.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

- FEC is to insert redundant data into normal packets so that the line has certain error tolerance. Some bandwidth, however, must be consumed. Enabling FEC enhances the error correction capability of the line but at the same time occupies certain bandwidth. Determine whether to enable FEC according to the actual line planning.
- If a large number of ONTs are online, enabling the FEC function of the optical port may cause some ONTs to go offline. Therefore, it is recommended that you do not configure the FEC function of the optical port when there are online ONTs.

Step 3 Configure the renewal time of the ONT key.

Run the **port portid ont-password-renew** command to configure the interval for renewing the ONT key. To ensure the system security, the key renewal must be configured.

Step 4 Configure the compensation distance in the ranging.

Run the **port portid range** command to configure the compensation distance range of the GPON port ranging.

Step 5 (Optional) Configure the DBA calculation period on a GPON port basis.

When different GPON ports provide different access services, the bandwidth delays on these ports are different. In this case, the DBA calculation period needs to be configured on a GPON port basis.

1. In GPON board mode, run the **port dba bandwidth-assignment-mode** command to configure the DBA mode on a GPON port.
2. In diagnose mode, run the **gpon port dba calculate-period** command to configure the DBA calculation period on the GPON port.

NOTE

- The DBA calculation period on a GPON port can be configured only when the DBA mode is set to **manual** on this GPON port.
- By default, the DBA mode on a GPON port is **default**, which means the global DBA mode is used as the bandwidth assignment mode for the GPON port. In this case, if the global DBA mode is modified by running the **gpon dba bandwidth-assignment-mode** command, the bandwidth assignment mode on the GPON port is also modified. If the DBA mode on a GPON port is not **default**, the bandwidth assignment mode on the GPON port is not affected by the global DBA mode.
- If ONTs are configured on a GPON port, modifying the DBA mode is not allowed on this GPON port.
- For the TDM service, the DBA mode must be set to **min-loop-delay**.

----End

Example

To enable the FEC function of GPON port 0/1/0, set the key renewal interval for the ONT connected to this port to 10 h, the minimum ranging compensation distance to 10 km, and the maximum ranging compensation distance to 15 km, do as follows:

```
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#port 0 fec enable
huawei(config-if-gpon-0/1)#port 0 ont-password-renew 10
huawei(config-if-gpon-0/1)#port 0 range min-distance 10 max-distance 15
```

 Copyright © Huawei Technologies Co., Ltd.

```
This command will result in the ONT's re-register in the port.
Are you sure to execute this command? (y/n)[n]: y
```

To set the global DBA mode to **min-loop-delay**, DBA mode on GPON port 0/1/0 to **manual**, and DBA calculation period to **4**, do as follows:

```
huawei(config)#gpon dba bandwidth-assignment-mode min-loop-delay
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#port dba bandwidth-assignment-mode 0 manual
huawei(config-if-gpon-0/1)#quit
huawei(config)#diagnose
huawei(diagnose)%%gpon port dba calculate-period 0/1/0 4
```

2.9 GPON Configuration Guide (Distributed mode)

GPON configurations include the configurations on GPON profiles, ONTs, and ports. The following section describes configuration methods.

Context

The xPON mode includes two types: distributed (discrete) mode and profile mode. The differences between the two modes are as follows:

- In the distributing mode, ONTs cannot be added in batches. Instead, ONTs need to be configured one by one.
- In profile mode, you can pre-configure ONT line profiles and ONT service profiles and bind ONTs of the same configurations to the same profile to add them in batches, which significantly improves service provisioning efficiency.

The xPON mode is determined during site provisioning and will not be changed.

You can run the **display xpon mode** command to query the xPON mode of the current system.

2.9.1 Configuring a GPON ONT Profile (Distributed Mode)

In distributed mode, GPON ONT profiles include the GPON ONT capability profile and the GPON ONT alarm profile. This topic describes how to configure these profiles.

Context

GPON ONT profiles contain the parameters required for configuring the GPON access service, of which,

- DBA profiles specify GPON traffic parameters. The DBA profile bound to an OLT enables the OLT to dynamically allocate bandwidths, improving upstream bandwidth utilization.
- In distributed mode, the GPON ONT capability profile contains the physical port type and quantity of the ONU, mapping mode from service port to GEM port, and traffic control type.
- The GPON ONT alarm profile provides a series of alarm threshold parameters that are used for performance measurement and monitoring of activated ONU lines. After a GPON alarm profile is bound to an ONU, the ONU sends alarms to the log host and the NMS if the performance statistics of the line exceed the threshold that is specified in the profile.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

In this document, ONUs include MDUs and ONTs.

2.9.1.1 Configuring a DBA Profile

A DBA profile defines the traffic parameters of xPON and can be bound to a T-CONT to dynamically allocate the bandwidth and improve the usage of the upstream bandwidth.

Default Configuration

Table 2-12 lists the default settings of the DBA profiles.

Table 2-12 Default settings of the DBA profiles

Parameter	Default Setting	Remarks
Default DBA profile ID in the system	0-9	You can run the **display dba-profile all** command to query the parameter values of each default DBA profile.

Procedure

Step 1 Add a DBA profile.

Run the **dba-profile add** command to add a DBA profile. By default, the system provides DBA profiles 0–9 with typical traffic parameter values. The default profiles cannot be added or deleted.

NOTE

- By default, a T-CONT is not bound to any DBA profile. You must bind a T-CONT to a DBA profile.
- When you add a DBA profile, the bandwidth value must be a multiple of 64. If you enter a bandwidth value not of a multiple of 64, the system adopts the closest multiple of 64 that is smaller than the value you enter.

Step 2 Query a DBA profile.

Run the **display dba-profile** command to query a DBA profile.

----End

Example

To add a DBA profile with the type of type3, set the profile name to DBA_100M and the required bandwidth to 100 Mbit/s, do as follows:

```
huawei(config)#dba-profile add profile-name DBA_100M type3 assure 102400 max 102400
huawei(config)#display dba-profile profile-name DBA_100M
```

 Copyright © Huawei Technologies Co., Ltd.

2.9.1.2 Configuring a GPON ONT Capacity Profile (Distributed Mode)

A GPON ONT capability profile identifies the actual capability of a GPON ONU. After an ONT is added and bound to a GPON ONT capability profile, the ONU carries the corresponding services according to parameters configured in the capability profile.

Context

- All GPON ONUs must be bound to the GPON ONT capability profile. Specify the ONT capability profile when running the **ont add** command to add an ONU offline or running the **ont confirm** command to confirm an automatically discovered ONU.
- Currently, the system provides seven default ONT capability profiles that are solidified in the system. The default profiles cannot be modified. The default profile IDs range from 1-7. The reserved ONT capability profile IDs are 8-16.
- The contents of the capability profile restrict the port number that is used in commands for GEM port mapping, T-CONT/PQ mapping, and the ONT VLAN management.
- The ONT capability profile must be configured according to the actual capability of the ONU. Different the capability profile parameters vary according to different ONUs.

Procedure

Step 1 Run the **ont-profile add** command to configure an ONT capability profile.

- When you add an ONT capability profile, if the profile ID is not specified, the system automatically allocates the least idle profile ID; if the profile name is not specified, the system adopts the default name **ont-profile_x**, where, x is the corresponding ONT capability profile ID.
- The system supports up to 128 ONT capability profiles.
- The system default profiles include the MDU profile and several common ONT (such as OT925, HG850, and HG810) profiles, which can be directly used. It is recommended to manually configure an ONT capability profile only when the default ONT capability profile fails to meet actual requirements.
- When you add an MDU profile manually, the number of the ports must be set to zero.

Step 2 Run the **display ont-profile** command to query the ONT capability profile.

----End

Example

To add an ONT capability profile for ONT HG850a with the profile index of 30, 2 POTS ports, 4 Ethernet ports, mapping mode of VLANID, and flow control type of PQ, and query the capability profile after the configuration is complete, do as follows:

```
huawei(config)#ont-profile add profile-id 30
{ <cr>|profile-name<K> }:

  Command:
```

 Copyright © Huawei Technologies Co., Ltd.

```
        ont-profile add profile-id 30
  Press 'Q' or 'q' to quit input
>  Are you sure you want to set the number of POTS ports to auto-adaptive? (y/n)
 [n]:
>  Number of POTS ports<0-8> [0]:2
>  Are you sure you want to set the number of ETH ports to auto-adaptive? (y/n)
[n]:
>  Number of ETH ports<0-8>  [0]:4
>  Are you sure you want to set the number of VDSL ports to auto-adaptive? (y/n)
[n]: y
>  TDM port type<1-E1,2-T1> [1]:
>  TDM service type<1-TDMoGEM> [1]:
>  Number of TDM ports<0-8> [0]:
>  Number of MOCA ports<0-8> [0]:
>  Are you sure you want to set the number of CATV UNI ports to auto-adaptive? (
y/n) [n]:
>  Number of CATV UNI ports<0-8> [0]:
>  Mapping mode<1-VLANID, 2-802_1pPRI, 3-VLANID_802_1pPRI, 4-PORTID,
   5-PORTID_VLANID, 6-PORTID_802_1pPRI, 7-PORTID_VLANID_802_1pPRI,
   9-IPTOS, 10-VLANID_IPTOS> [1]:
>  The type of flow control<1-PQ, 2-GEMPORT-CAR, 3-FLOW-CAR> [1]:
  Adding an ONT profile succeeded
  Profile ID  : 30
  Profile name: ont-profile_30
huawei(config)#display ont-profile profile-id 30
  ------------------------------------------------------------------------------
  Profile ID  : 30
  Profile name: ont-profile_30
  ------------------------------------------------------------------------------
  Number of POTS ports:              2
  Number of ETH ports:               4
  Number of VDSL ports:              0
  TDM port type:                    E1
  TDM service type:                 TDMoGem
  Number of TDM ports:               0
  Number of MOCA ports:              0
  Number of CATV UNI ports:          0
  Mapping mode:                     VLAN ID
  The type of flow control:         PQ
  ------------------------------------------------------------------------------
  Binding times:                    0
  ------------------------------------------------------------------------------
```

2.9.1.3 Configuring a GPON ONT Alarm Profile

This topic describes how to add an alarm profile, and configure most of the performance parameters for various ONT lines as a profile. After the alarm profile is configured and bound successfully, the ONT can directly use the profile when it is activated.

Context

An ONT alarm profile defines a series of alarm thresholds that are used to monitor the performance of an activated ONT line. When the statistics result of a parameter reaches the alarm threshold, the NE is notified and an alarm is sent to the log server and the NMS.

The system has a default alarm threshold profile with ID 1. This profile cannot be deleted but can be modified.

Procedure

Step 1 Run the **gpon alarm-profile add** command to add a GPON ONT alarm profile.

 Copyright © Huawei Technologies Co., Ltd.

If default alarm thresholds are used in a new alarm profile, the system does not report any alarm.

Step 2 Run the **display gpon alarm-profile** command to query the alarm profile.

----End

Example

To add GPON ONT alarm profile 5, set the alarm threshold for the packet loss of the GEM port to 10, set the alarm threshold for the number of mis-transmitted packets to 30, and use the default value 0 for all other thresholds, do as follows:

```
huawei(config)#gpon alarm-profile add profile-id 5
{ <cr>|profile-name<K> }:

  Command:
        gpon alarm-profile add profile-id 5
  Press 'Q' or 'q' to quit input
>  GEM port loss of packets threshold (0~100)[0]:10
>  GEM port misinserted packets threshold (0~100)[0]:30
>  GEM port impaired blocks threshold (0~100)0[0]:
>  Ethernet FCS errors threshold (0~100)[0]:
>  Ethernet excessive collision count threshold (0~100)[0]:
>  Ethernet late collision count threshold (0~100)[0]:
>  Too long Ethernet frames threshold (0~100)[0]:
>  Ethernet buffer (Rx) overflows threshold (0~100)[0]:
>  Ethernet buffer (Tx) overflows threshold (0~100)[0]:
>  Ethernet single collision frame count threshold (0~100)[0]:
>  Ethernet multiple collisions frame count threshold (0~100)[0]:
>  Ethernet SQE count threshold (0~100)[0]:
>  Ethernet deferred transmission count threshold (0~100)[0]:
>  Ethernet internal MAC Tx errors threshold (0~100)[0]:
>  Ethernet carrier sense errors threshold (0~100)[0]:
>  Ethernet alignment errors threshold (0~100)[0]:
>  Ethernet internal MAC Rx errors threshold (0~100)[0]:
>  PPPOE filtered frames threshold (0~100)[0]:
>  MAC bridge port discarded frames due to delay threshold (0~100)[0]:
>  MAC bridge port MTU exceeded discard frames threshold (0~100)[0]:
>  MAC bridge port received incorrect frames threshold (0~100)[0]:
>  CES general error time threshold(0~100)[0]:
>  CES severely time threshold(0~100)[0]:
>  CES bursty time threshold(0~100)[0]:
>  CES controlled slip threshold(0~100)[0]:
>  CES unavailable time threshold(0~100)[0]:
>  Drop events threshold(0~100)[0]:
>  Undersize packets threshold(0~100)[0]:
>  Fragments threshold(0~100)[0]:
>  Jabbers threshold(0~100)[0]:
>  Failed signal of ONT threshold(Format:1e-x, x: 3~8)[3]:
>  Degraded signal of ONT threshold(Format:1e-x, x: 4~9)[4]:
>  FEC uncorrectable code words threshold(0~1101600000)[0]:
>  FEC correctable code words threshold(0~1101600000)[0]:
>  Upstream PQ discarded byte alarm threshold(0~65535)[0]:
>  Downstream PQ discarded byte alarm threshold(0~65535)[0]:
>  Encryption key errors threshold(0~100)[0]:
>  XGEM key errors threshold(0~100)[0]:
>  XGEM HEC error count threshold(0~100)[0]:

  Adding an alarm profile succeeded
  Profile ID  : 5
  Profile name: alarm-profile_5
```

 Copyright © Huawei Technologies Co., Ltd.

2.9.2 Configuring a GPON ONT (Distributed Mode)

The OLT provides end users with services through the ONT. The OLT can manage the ONT and the ONT can work in the normal state only after the channel between the OLT and the ONT is available.

Prerequisites

The GPON ONT profile is already created. **Configuring a GPON ONT Capability Profile** and **Configuring a GPON ONT Alarm Profile** are already completed.

Context

The OLT uses the ONT Management and Control Interface (OMCI) protocol to manage and configure the GPON ONT, and supports the offline configuration of the ONT. The ONT does not need to save the configuration information locally. This helps to provision services.

Table 2-13 lists the default settings of the GPON ONT.

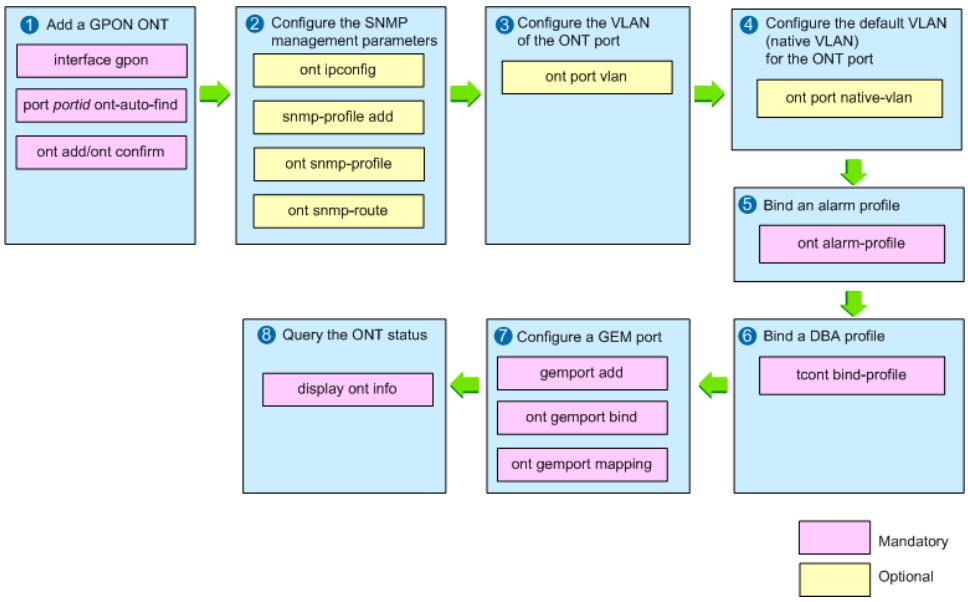
Table 2-13 Default settings of the GPON ONT

Parameter	Default Setting
ONT auto-find function of a GPON port	Disabled
ONT status after an ONT is added	Activated
Default VLAN of the ONT port	1

Configuration Process

Figure 2-42 shows the process of configuring a GPON ONT.

Figure 2-42 Process of configuring a GPON ONT



 Copyright © Huawei Technologies Co., Ltd.

Procedure

Step 1 Add a GPON ONT.

1. Run the **interface gpon** command to enter the GPON mode.
2. Run the **port portid ont-auto-find** command to enable the auto-find function of the ONT. After the function is enabled, you can add an ONT according to the information reported by the system. By default, the ONT auto-find function of a GPON port is disabled.

   NOTE

   An auto-find ONT is in the auto-find state. The auto-find ONT can work in the normal state only after it is confirmed or added.
3. Run the **ont add** command to add an ONT offline, or run the **ont confirm** command to confirm the auto-find ONT.

   When ONTs are added or confirmed, the system provides four authentication modes: SN, password, SN+password, LOID+CHECKCODE.

   - SN authentication: The OLT detects the serial number (SN) reported by an ONT. If the SN is consistent with the OLT configuration, authentication is passed and the ONT goes online. This mode requires recording all ONT SNs. Hence, it is used to confirm auto discovery ONTs and is not applicable to adding ONTs in batches.
   - Password authentication: The OLT detects the password reported by an ONT. If the password is consistent with the OLT configuration, the ONT goes online normally. This mode requires planning ONT passwords and does not require manually recording ONT SNs. Hence, it is applicable to adding ONTs in batches. The password authentication provides two discovery modes: always-on and once-on.
     - always-on: After first password authentication is passed, no SN is allocated and password authentication is always used in subsequent authentications. This discovery mode is easy for future maintenance. In the always-on discovery mode, configurations do not need to be modified when an ONT is replaced and only the password is required. The always-on discovery mode has lower security. If other users know the password, the users will illegally have service permissions.
     - once-on: Password authentication is used for the first authentication. After the authentication is successful, the SN is automatically matched. The SN+password authentication mode is used for subsequent authentications. The ONT can go online only when both the password and SN are correct. The once-on authentication mode has high security. After an ONT is replaced or the password is mistakenly changed, the ONT needs to run the **ont re-discovery** command to be re-discovered, which requires more maintenance effort.
   - SN+password authentication: The OLT detects the password and SN reported by an ONT. If the password and SN are consistent with the OLT configuration, the ONT goes online normally. This authentication mode has the highest security but it requires manually recording ONT SNs.

 Copyright © Huawei Technologies Co., Ltd.

- LOID+CHECKCODE authentication: defined by a telecom operator. In this authentication mode, LOID has 24 bytes, and CHECKCODE has 12 bytes and is optional. Whether 24 bytes or 36 bytes are used for authentication depends on data planning, which is unified over the entire network. The OLT determines whether LOID+CHECKCODE reported by the ONT is the same as the configured one. If they are the same, the ONT authentication is passed. If they are different, the OLT obtains the ONT password and compares it with the last 10 bytes of the LOID. If they are the same, the ONT authentication is also passed. This operation is for compatibility with the ONTs using password authentication.

Adding ONTs in offline mode is applicable to the batch deployment scenario. All ONTs are added to the OLT to complete service provisioning beforehand. When a use subscribes to the service, an installation engineer takes an ONT to the user's house and completes configurations. After the ONT goes online and passes authentication, the service is provisioned. Generally, the password authentication mode or LOID authentication mode is used to confirm an ONT.

Adding ONTs in auto discovery mode is applicable to the scenario where a small number of ONTs are added. When users subscribe to the service, installation engineers take ONTs to the users' houses. After the ONTs go online, the OLT confirms the ONTs one by one. Generally, the MAC address authentication mode is used to confirm the ONTs.

NOTE

- If the ONU is an independent NE and is directly managed by the NMS through the SNMP management mode, select the SNMP management mode. For this mode, you only need to configure the parameters for the GPON line and the parameters for the management channel on the OLT.
- If the ONU is not an independent NE and all its configuration data is issued by the OLT through OMCI, select the OMCI management mode. For this mode, you need to configure all parameters (including line parameters, UNI port parameters, and service parameters) that are required for the ONU on the OLT.
- Generally, the ONT management mode is set to the OMCI mode.

Step 2 (Optional) Configure the SNMP management parameters. When the ONT management mode is the SNMP mode, you need to configure the SNMP management parameters for the ONT. The procedure is as follows:

1. Run the **ont ipconfig** command to configure the management IP address of the ONT. The IP address should not be in the same subnet for the IP address of the VLAN port.
2. If an ONU is not managed as an independent NE, all configurations of the ONU are managed by the OLT through the OMCI protocol, and the OMCI management mode is selected. In this mode, you need to configure all the parameters (including line parameters, UNI port parameters, and service parameters) required by the ONU on the OLT.
3. Run the **ont snmp-route** command to configure a static route for the NMS server, that is, configure the IP address of the next hop.

Step 3 (Optional) Configure the VLAN of the ONT port.

Run the **ont port vlan** command to configure the VLAN of the ONT port.

Step 4 (Optional) Configure the default VLAN (native VLAN) for the ONT port.

 Copyright © Huawei Technologies Co., Ltd.

Run the **ont port native-vlan** command to configure the default VLAN for the ONT port.

- If the packets reported from a user (such a PC) to the ONT are untagged, the packets are tagged with the default VLAN of the port on the ONT and then reported to the OLT.
- If the packets reported from a user to the ONT are tagged, you need to configure the port VLAN of the ONT to be the same as the VLAN in the user tag. The packets are not tagged with the default VLAN of the port on the ONT but are reported to the OLT with the user tag.

Step 5 Bind an alarm profile.

Run the **ont alarm-profile** command to bind an alarm profile. Ensure that **2.8.1.4 Configuring a GPON ONT Alarm Profile** is completed before the configuration.

Step 6 Bind a DBA profile.

Run the **tcont bind-profile** command to bind a DBA profile to a T-CONT.

A DBA profile can be bound to a T-CONT after an ONT is added.

Step 7 Configure a GEM port.

1. Run the **gemport add** command to add a GEM port. When adding a GEM port, select the correct attribute according to the service type.
2. Run the **ont gemport bind** command to bind the GEM port to an ONT T-CONT, that is, allocating the T-CONT resources to the GEM port.

   NOTE

   If traffic streams are configured on a GEM port and an ONT is the working ONT in a single-homing protection group, the GEM port cannot be bound to or unbound from the ONT.
3. Run the **ont gemport mapping** command to create the mapping between the GEM port and the ONT-side service.

Step 8 Query the ONT status.

Run the **display ont info** command to query the ONT running status, configuration status, and matching status.

----End

Example

To add five ONTs in offline mode with password authentication mode (ONT passwords are 0100000001-0100000005), set the discovery mode of password authentication to always-on, and bind ONT capability profile 30, do as follows:

```
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#ont add 0 password-auth 0100000001 always-on profile-id 30 manage-mode omci
huawei(config-if-gpon-0/1)#ont add 0 password-auth 0100000002 always-on profile-id 30 manage-mode omci
huawei(config-if-gpon-0/1)#ont add 0 password-auth 0100000003 always-on profile-id 30 manage-mode omci
huawei(config-if-gpon-0/1)#ont add 0 password-auth 0100000004 always-on profile-id 30 manage-mode omci
huawei(config-if-gpon-0/1)#ont add 0 password-auth 0100000005 always-on profile-id 30 manage-mode omci
```

 Copyright © Huawei Technologies Co., Ltd.

To add an ONT that is managed by the OLT through the OMCI protocol, confirm this ONT according to the SN 3230313185885B41 automatically reported by the system, and bind the ONT with capability profile 3 that match the ONT, do as follows:

```
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#port 0 ont-auto-find enable
huawei(config-if-gpon-0/1)#ont confirm 0 sn-auth 3230313185885B41 profile-id 3
 manage-mode omci
```

To add an ONU that is managed as an independent NE and whose SN is known as 3230313185885641, bind the ONU with capability profile 4 that matches the ONU, configure the NMS parameters for the ONU, and set the management VLAN to 100, do as follows:

```
huawei(config)#snmp-profile add profile-id 1 v3 public private 10.10.5.53 161 bodyname
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#ont add 0 2 sn-auth 3230313185885641 profile-id 4 manage-mode snmp
huawei(config-if-gpon-0/1)#ont ipconfig 0 2 static ip-address 10.20.20.20 mask 255.255.255.0 gateway
10.10.20.1 vlan 100
huawei(config-if-gpon-0/1)#ont snmp-profile 0 2 profile-id 1
huawei(config-if-gpon-0/1)#ont snmp-route 0 2 ip-address 10.10.20.190 mask
255.255.255.0 next-hop 10.10.20.100
```

2.9.3 Configuring a GPON Port

To work normally and carry the service, a GPON port must be enabled first. This topic describes how to enable a GPON port and configure related attributes of the port.

Default Configuration

Table 2-14 lists the default settings of the GPON port.

Table 2-14 Default settings of the GPON port

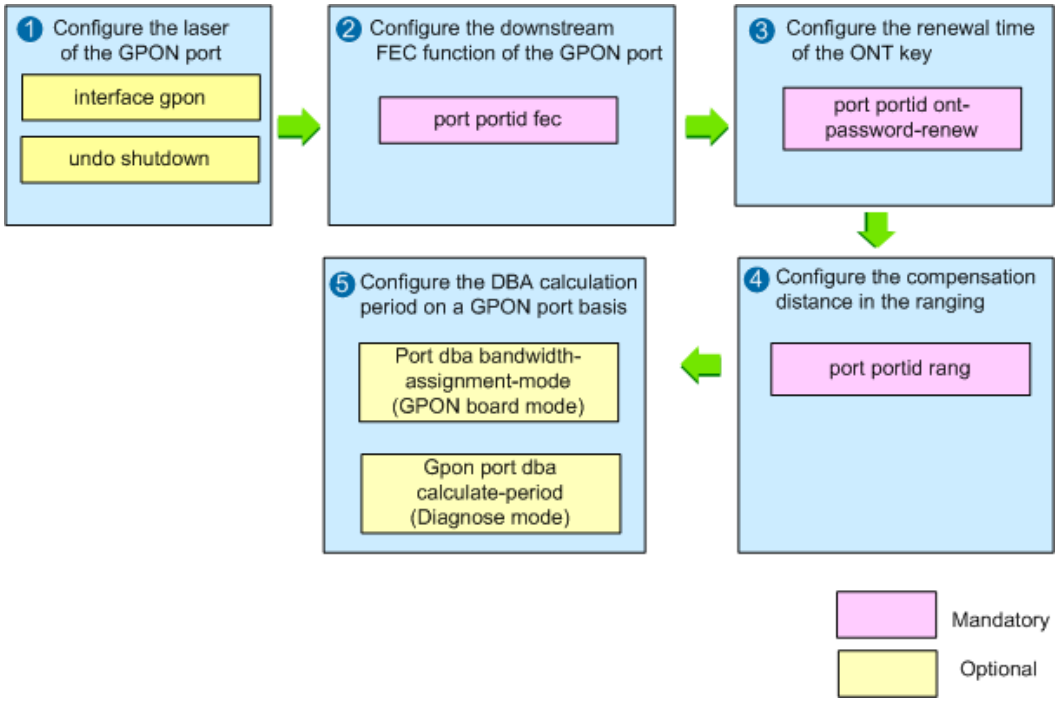
Parameter	Default Setting
GPON port	Enabled
Downstream FEC function of the GPON port	Disabled
Compensation distance range of the GPON port ranging	Minimum logical distance: 0 km; maximum logical distance: 20 km

Configuration Process

Figure 2-43 shows the process of configuring a GPON Port.

 Copyright © Huawei Technologies Co., Ltd.

Figure 2-43 Process of configuring a GPON Port



Procedure

Step 1 (Optional) Configure the optical port switch.

1. Run the **interface gpon** command to enter GPON mode.
2. Run the **undo shutdown** command to enable the laser of the GPON port. By default, the laser of the GPON port is enabled and the GPON port is available. In this case, skip this step.

NOTICE

- If an optical port is not required, run the **shutdown** command to disable the optical port.
- Disabling a PON port that carries services will cause the interruption of such services.

Step 2 Enables or disables the downstream FEC function of a port.

Run the **port portid fec** command to configure the FEC function of the GPON port. By default, the function is disabled.

NOTE

- FEC is to insert redundant data into normal packets so that the line has certain error tolerance. Some bandwidth, however, must be consumed. Enabling FEC enhances the error correction capability of the line but at the same time occupies certain bandwidth. Determine whether to enable FEC according to the actual line planning.
- If a large number of ONTs are online, enabling the FEC function of the optical port may cause some ONTs to go offline. Therefore, it is recommended that you do not configure the FEC function of the optical port when there are online ONTs.

 Copyright © Huawei Technologies Co., Ltd.

Step 3 Configure the renewal time of the ONT key.

Run the **port portid ont-password-renew** command to configure the interval for renewing the ONT key. To ensure the system security, the key renewal must be configured.

Step 4 Configure the compensation distance in the ranging.

Run the **port portid range** command to configure the compensation distance range of the GPON port ranging.

Step 5 (Optional) Configure the DBA calculation period on a GPON port basis.

When different GPON ports provide different access services, the bandwidth delays on these ports are different. In this case, the DBA calculation period needs to be configured on a GPON port basis.

1. In GPON board mode, run the **port dba bandwidth-assignment-mode** command to configure the DBA mode on a GPON port.
2. In diagnose mode, run the **gpon port dba calculate-period** command to configure the DBA calculation period on the GPON port.

NOTE

- The DBA calculation period on a GPON port can be configured only when the DBA mode is set to **manual** on this GPON port.
- By default, the DBA mode on a GPON port is **default**, which means the global DBA mode is used as the bandwidth assignment mode for the GPON port. In this case, if the global DBA mode is modified by running the **gpon dba bandwidth-assignment-mode** command, the bandwidth assignment mode on the GPON port is also modified. If the DBA mode on a GPON port is not **default**, the bandwidth assignment mode on the GPON port is not affected by the global DBA mode.
- If ONTs are configured on a GPON port, modifying the DBA mode is not allowed on this GPON port.
- For the TDM service, the DBA mode must be set to **min-loop-delay**.

----End

Example

To enable the FEC function of GPON port 0/1/0, set the key renewal interval for the ONT connected to this port to 10 h, the minimum ranging compensation distance to 10 km, and the maximum ranging compensation distance to 15 km, do as follows:

```
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#port 0 fec enable
huawei(config-if-gpon-0/1)#port 0 ont-password-renew 10
huawei(config-if-gpon-0/1)#port 0 range min-distance 10 max-distance 15
  This command will result in the ONT's re-register in the port.
  Are you sure to execute this command? (y/n)[n]: y
```

To set the global DBA mode to **min-loop-delay**, DBA mode on GPON port 0/1/0 to **manual**, and DBA calculation period to **4**, do as follows:

```
huawei(config)#gpon dba bandwidth-assignment-mode min-loop-delay
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#port dba bandwidth-assignment-mode 0 manual
huawei(config-if-gpon-0/1)#quit
huawei(config)#diagnose
huawei(diagnose)%%gpon port dba calculate-period 0/1/0 4
```

 Copyright © Huawei Technologies Co., Ltd.

2.10 GPON Reference Standards and Protocols

The reference standards and protocols of the GPON feature are as follows:

Standard No.	Description
ITU-T G.984.1	General Characteristics. This protocol mainly describes the basic features and major protection modes of GPON.
ITU-T G.984.2	Physical Media Dependent (PMD) Layer Specification. This protocol mainly describes the PMD layer parameters, including physical parameters (such as the transmit optical power, receiver sensitivity, and overload optical power) of optical transceivers, and also defines optical budget of different levels, for example, the most common Class B+.
ITU-T G.984.3	Transmission Convergence Layer Specification. This protocol mainly describes the TC layer specifications, including the upstream and downstream frame structures and GPON principle.
ITU-T G.984.4	ONT Management And Control Interface Specification. This protocol mainly describes the GPON management and maintenance protocols, such as OAM, PLOAM, and OMCI.
ITU-T G.984.5	Enhancement Band. This protocol mainly describes the GPON wavelength planning, including reserving bands for next-generation PON.
ITU-T G.984.7	Long Reach. This protocol mainly describes several long reach PON schemes for extending GPON transmission distance.
ITU-T G.988	ONU management and control interface (OMCI) specification.
TR-156	Using GPON Access in the context of TR-101.

 Copyright © Huawei Technologies Co., Ltd.

3 XG(S)-PON

XG(S)-PON is evolved from the existing GPON technology, XG(S)-PON technologies include XG-PON and XGS-PON.

3.1 Introduction to XG(S)-PON

3.2 XG(S)-PON Network Structure

3.3 XG(S)-PON Working Principle

3.4 XG(S)-PON Key Technologies

3.5 XG(S)-PON Security Technologies

3.6 Evolution From GPON to XG(S)-PON

3.7 XG(S)-PON Reference Standards and Protocols

3.1 Introduction to XG(S)-PON

What Is XG(S)-PON?

XG(S)-PON is evolved from the existing GPON technology, the factors promoting the technology evolution to XG(S)-PON are as follows:

- Developing services require higher bandwidths and the GPON technology cannot meet bandwidth requirements.
- Innovative access technologies on the user side require higher bandwidths and the GPON technology will face bandwidth bottlenecks.
- A greater split ratio and a longer transmission distance increase network construction investments.

Specification Differences Between GPON, XG-PON and XGS-PON

The following table lists specification differences between the two technologies.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

Here only describes the technical specification of the GPON, XG-PON and XGS-PON, does not represent the current support capabilities of the product, the product capabilities please refer to the specific specifications.

Specification	GPON	XG(S)-PON	
		XG-PON	XGS-PON
Wavelength	• Downstream: 1480-1500 nm • Upstream: 1290-1330 nm	• Downstream: 1575-1580 nm • Upstream: 1260-1280 nm	• Downstream: 1575-1580 nm • Upstream: 1260-1280 nm
Center wavelength	• Downstream: 1490 nm • Upstream: 1310 nm	• Downstream: 1577 nm • Upstream: 1270 nm	• Downstream: 1577 nm • Upstream: 1270 nm
Maximum Line Rate	• Downstream: 2.488 Gbit/s • Upstream: 1.244 Gbit/s	• Downstream: 9.953 Gbit/s • Upstream: 2.488 Gbit/s	• Downstream: 9.953 Gbit/s • Upstream: 9.953 Gbit/s
Frame Structure	GEM	XGEM	XGEM

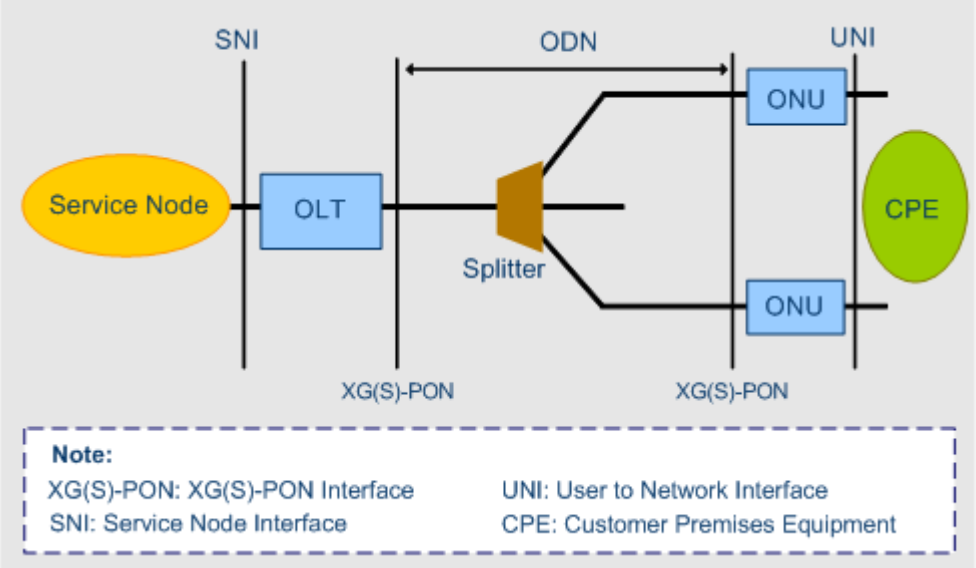
3.2 XG(S)-PON Network Structure

Network Structure

A XG(S)-PON network is of the point-to-multipoint (P2MP) type, which is the same as that of a GPON network. **Figure 3-1** shows a XG(S)-PON network structure.

Copyright © Huawei Technologies Co., Ltd.

Figure 3-1 Network Structure



The XG(S)-PON network contains an optical line terminal (OLT), optical network units (ONUs), and an optical distribution network (ODN).

- The Optical line terminal (OLT) is an aggregation device located at the central office (CO) for terminating the PON protocol.
- Optical network units (ONUs) are located on the user side, providing various types of ports for connecting to user terminals.
- The Optical distribution network (ODN) is composed of passive optical components (POS) such as optical fibers, and one or more passive optical splitters. It interconnects the OLT and ONUs.

3.3 XG(S)-PON Working Principle

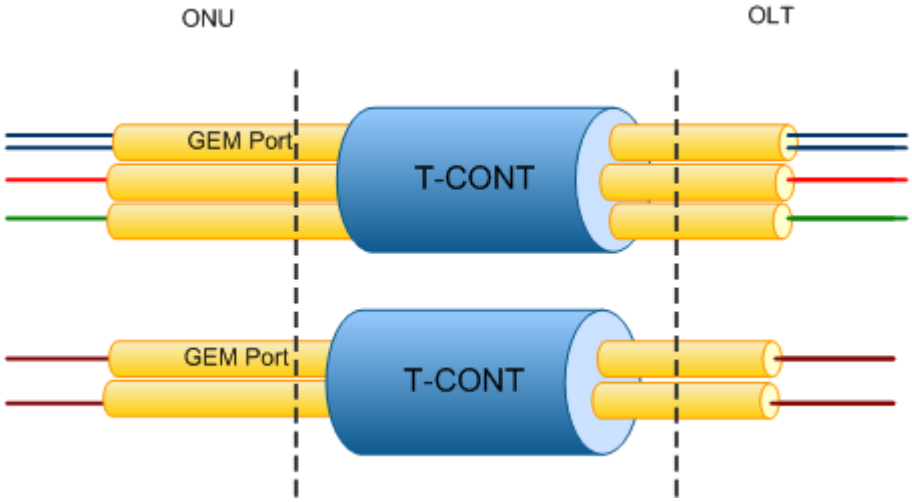
3.3.1 Basic Concepts

Service Multiplexing

GEM ports and T-CONTs divide a PON network into virtual connections for service multiplexing, as shown in **Figure 3-2**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 3-2 Working principles of service multiplexing



GEM Port

A GPON encapsulation mode (GEM) port is a virtual service channel that carries a service flow between the OLT and an ONU in an 10G PON system. The GEM port is similar to the virtual connection (identified by VPI/VCI) in asynchronous transfer mode (ATM). VPI is the acronym for virtual path identifier and VCI is the acronym for virtual channel identifier.

- Each GEM port is identified by a unique XGEM port ID.
- The XGEM port ID is globally allocated according to the 10G PON port by the OLT.
- A GEM port can carry one or more types of services.

T-CONT

A transmission container (T-CONT) is the basic control unit of upstream service flows in an 10G PON system, and is also the unit for carrying service flows in the upstream direction. All the GEM ports are mapped to T-CONTs, and the OLT uses dynamic bandwidth allocation (DBA) to schedule upstream transmission.

- A T-CONT can carry one or more GEM ports according to user configurations.
- A T-CONT is identified uniquely by Alloc-ID.
- The Alloc-ID is allocated according to the 10G PON port by the OLT.
- An ONU supports multiple T-CONTs configured for various service types.

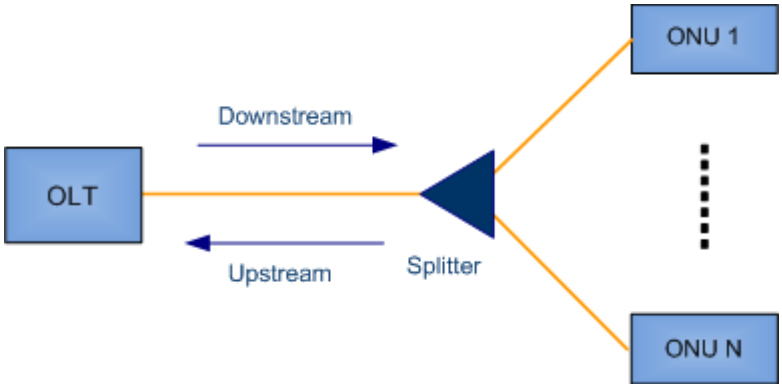
3.3.2 System Transmit Principles

System Transmit Principles

XG(S)-PON uses wavelength division multiplexing (WDM) to transmit data in different wavelengths on an ODN network. **Figure 3-3** shows the system transmit principles.

 Copyright © Huawei Technologies Co., Ltd.

Figure 3-3 System Transmit Principles



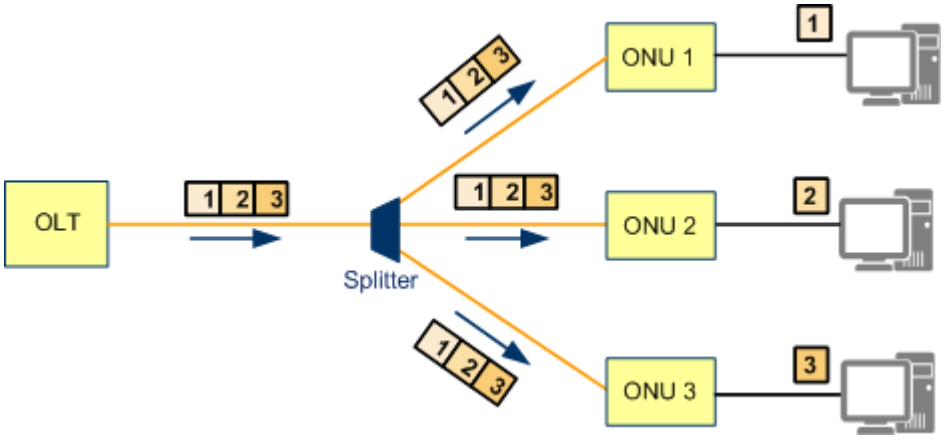
- Data is broadcast in the downstream direction.
- Data is transmitted in the TDMA mode (based on timeslots) in the upstream direction.

3.3.3 Working Principles of Downstream

Working Principles for Downstream Transmission

Figure 3-4 shows the XG(S)-PON working principles for downstream transmission.

Figure 3-4 Working principles for downstream transmission

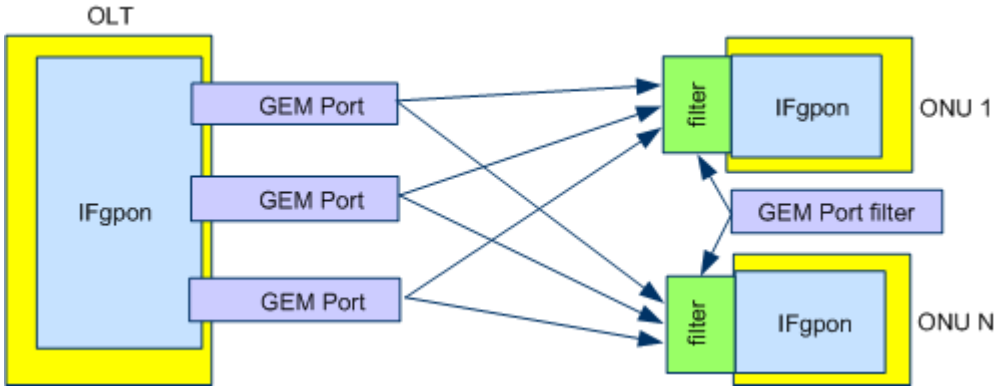


In the downstream direction, the OLT broadcasts data to all ONUs and the ONUs receive only desired data.

Copyright © Huawei Technologies Co., Ltd.

Data flow forwarding in the downstream direction

Figure 3-5 Data flow forwarding in the downstream direction



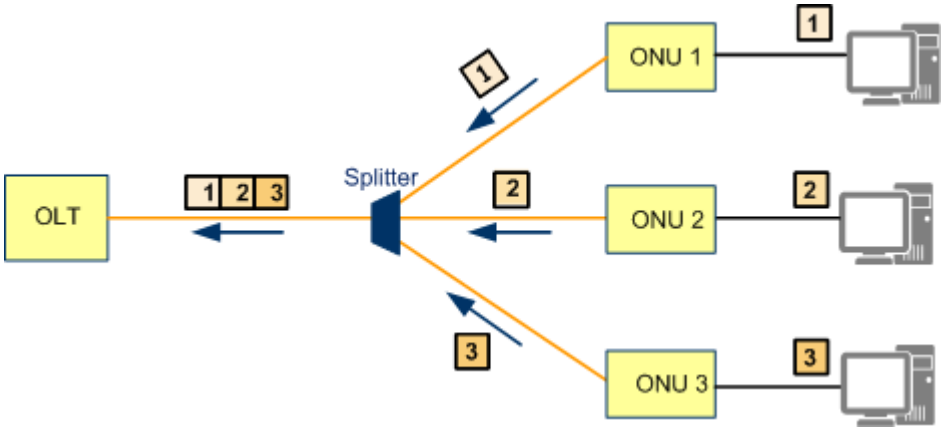
- On the OLT, data flows are encapsulated into GEM ports in service processing units.
- The OLT broadcasts the data to in the GEM ports to all ONUs.
- The ONU determines whether to process or discard the data according to the XGEM port ID.

3.3.4 Working Principle of Upstream

Working Principles for Upstream Transmission

Figure 3-6 shows the XG(S)-PON working principles for upstream transmission.

Figure 3-6 Working principles for upstream transmission

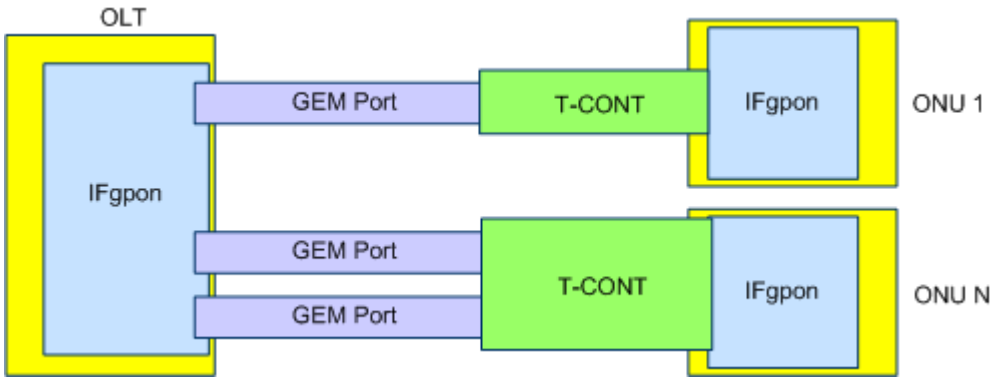


In the upstream direction, an ONU sends data to the OLT using an allocated timeslot. Such transmission ensures that all ONUs send data in a permitted sequence, which prevents upstream data collision.

 Copyright © Huawei Technologies Co., Ltd.

Data flow forwarding in the upstream direction

Figure 3-7 Data flow forwarding in the upstream direction



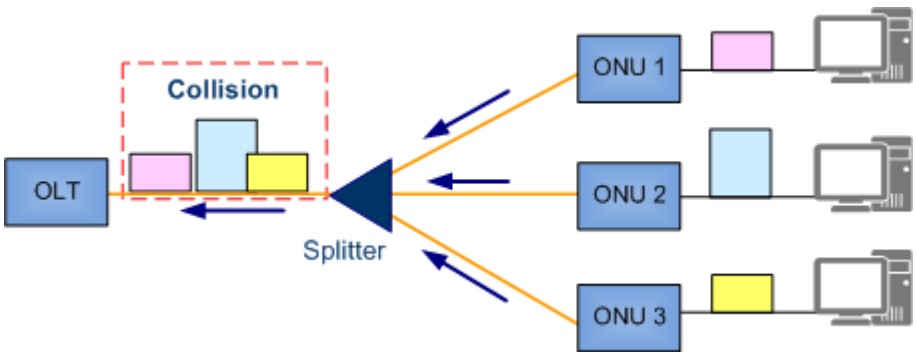
- On the ONU, data flows are encapsulated into GEM ports and mapped to transmission containers (T-CONTs).
- The OLT schedules data flows through the T-CONT, and the ONU sends data flows to the OLT according to the scheduling.
- The OLT decapsulates the data flows and sends them to service processing modules.

3.4 XG(S)-PON Key Technologies

3.4.1 Ranging

The logic reaches from optical network units (ONUs) to an optical line terminal (OLT) vary. The round trip delays (RTDs) between an OLT and ONUs also vary depending on time and environment. Therefore, collisions may occur when ONU sends data in TDMA mode (in this mode, only one of the ONUs connecting to a PON port sends data at a moment), as shown in **Figure 3-8**.

Figure 3-8 Cell transmission without ranging

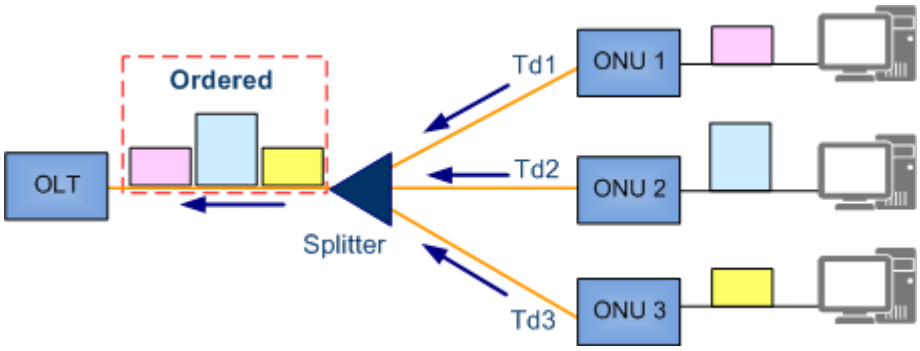


To prevent the collisions, ranging is enabled when an ONU registers for the first time. The OLT measures the RTD of each ONU in the ranging process and

Copyright © Huawei Technologies Co., Ltd.

calculates the equalization delay (EqD) of each ONU to ensure that the values of Teqd, which is equal to RTD plus EqD, of all ONUs connected to the same PON port are the same. Therefore, the logic reaches from ONUs to an OLT are the same, preventing collisions during upstream transmission.

Figure 3-9 Cell transmission with ranging



NOTE

In the ranging process, the OLT must open a window and pause upstream transmission channels of other ONUs.

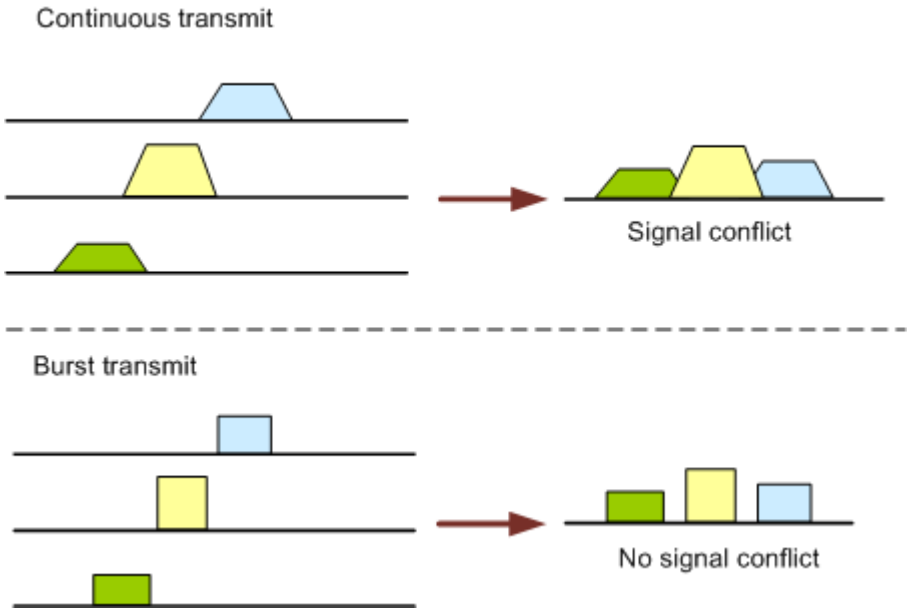
3.4.2 Burst Optical/Electrical Technology

In 10G PON upstream direction, Time Division Multiple Access (TDMA) is used. An optical network unit (ONU) transmits data only within the allocated timeslots. In the timeslots that are not allocated to it, the ONU disables the transmission of its optical transceiver to prevent other ONUs from being affected. The optical line terminal (OLT) then receives the upstream data from each ONU in a burst manner based on timeslots. Therefore, to ensure normal running of the 10G GPON system.

- **Figure 3-10** shows the burst transmit function supported by ONU-side optical modules.
- **Figure 3-11** shows the burst receive function supported by OLT-side optical modules.

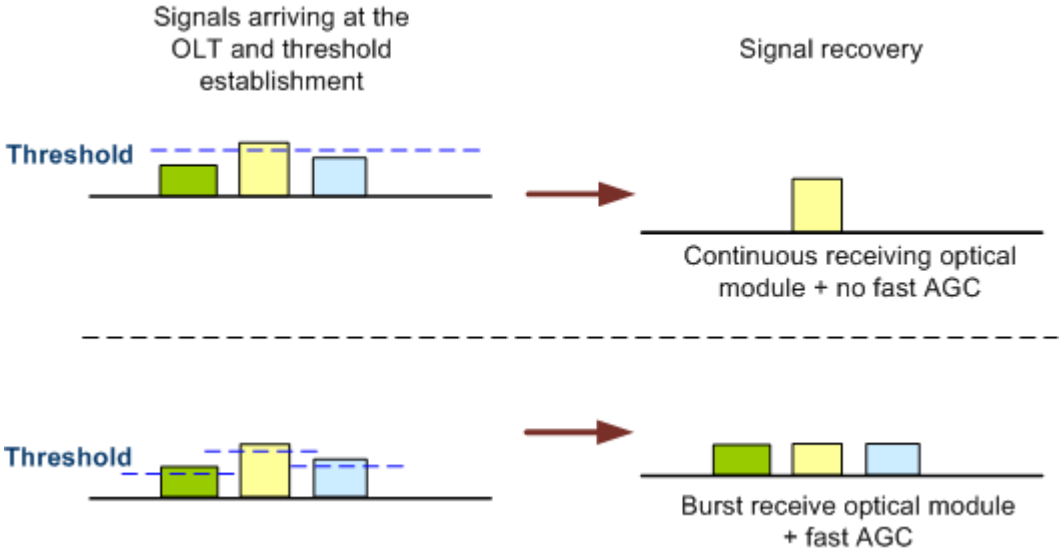
Copyright © Huawei Technologies Co., Ltd.

Figure 3-10 Burst transmit function supported by ONU-side optical modules



Ranging can be implemented to prevent cells transmitted by different ONUs from conflicting with each other on the OLT. However, the ranging accuracy is ± 1 bit and the cells transmitted by different ONUs have a protection time of several bits (not a multiple of 1 bit). If the ONU-side optical modules do not support the burst transmit function, the transmitted signals overlap and distortion occurs.

Figure 3-11 Burst receive function supported by OLT-side optical modules



- The distance from each ONU to the OLT varies and therefore the optical signal attenuation varies for each ONU. As a result, the power and level of packets received by an OLT at different timeslots various.
- If the OLT-side optical modules do not support the burst receive function, the OLT may restore incorrect signals because only the level greater than the threshold is considered valid and the signals with the level lower than the

 Copyright © Huawei Technologies Co., Ltd.

threshold cannot be restored. The automatic gain control (AGC) function dynamically adjusts the threshold of the receive optical power on the OLT based on the strength of the received optical signals to ensure that the signals of all ONUs can be restored.

NOTE

In the XG(S)-PON system, all data is broadcast downstream to ONUs. The transmission requires OLT-side optical modules to transmit optical signals continuously and ONU-side optical modules to receive optical signals continuously. Therefore, these optical modules are not required to support the burst receive and transmit function.

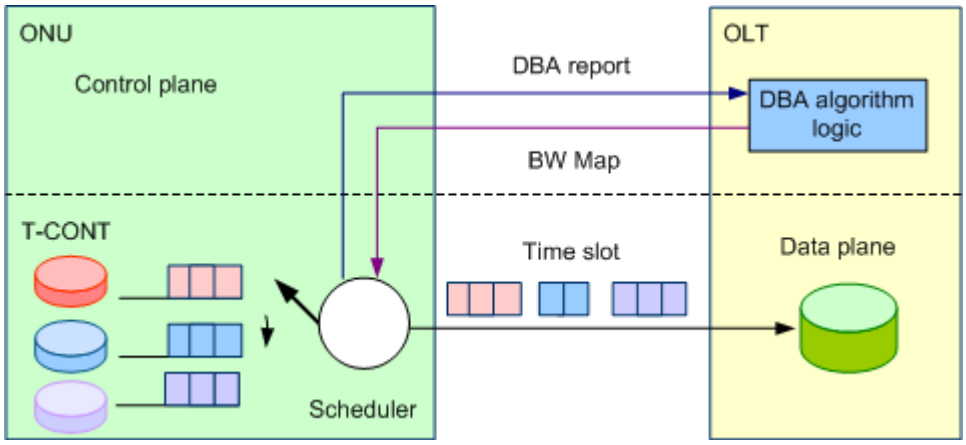
3.4.3 DBA

Working Principle

The OLT uses DBA to dynamically adjust the upstream bandwidth allocated to different ONUs to address the burst traffic on the ONUs, meeting the ONU upstream bandwidth requirements and improving the utilization of the PON upstream bandwidth.

Figure 3-12 shows the principles of DBA.

Figure 3-12 Principles of DBA



In the preceding figure:

- The DBA module in the OLT consistently collects DBA reports and uses the DBA algorithm to calculate the upstream bandwidth allocated to each ONU.
- The OLT sends the calculated result to each ONU using a bandwidth (BW) map.
- Each ONU transmits burst upstream data using permitted timeslots defined in the BW map.

Highlights and Applications

DBA offers the following features:

 Copyright © Huawei Technologies Co., Ltd.

- Based on ONUs' burst upstream service traffic, the OLT dynamically allocates an upstream bandwidth to each ONU in real time, improving upstream bandwidth utilization on PON ports.
- More users are supported on a PON port.
- Higher service bandwidths with burst requirements are supported than those before DBA is applied.

3.4.4 FEC

Context

Forward error correction (FEC) is mainly used for improving transmission quality of a line.

No ideal digital channel is available in practice. As a result, bit errors and jitter occur when digital signals are being transmitted over any transmission medium, deteriorating transmission quality on lines.

To resolve the problem, error correction mechanism is introduced.

- The mechanism can check and correct errors after data is transmitted to the peer end, for example, using FEC.
- The mechanism can check errors after data is transmitted to the peer end but not correct errors.

Working Principle

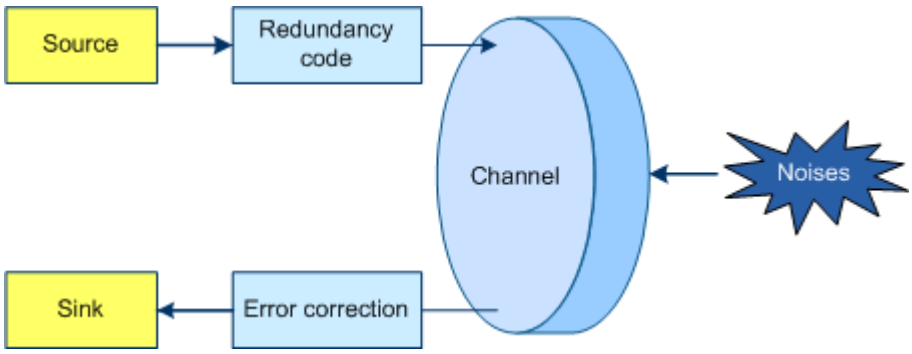
FEC is a data coding technology, which enables the RX end to check error bits in transmission based on the coding data.

Redundant codes are added to signals on the TX end. Then, the RX end checks the signals for errors based on error-correcting code (ECC) and corrects errors is there is any.

Common FEC codes include Hamming codes, RS codes, and convolutional codes.

Figure 3-13 shows FEC principles.

Figure 3-13 FEC principles



The FEC algorithm used by XG(S)-PON complies with ITU-T standards. Different FEC algorithms generate different bandwidth overheads.

Copyright © Huawei Technologies Co., Ltd.

Table 3-1 FEC algorithm used by XG(S)-PON

Mode	Upstream	Downstream
XG-PON	RS (248, 232)	RS (248, 216)
XGS-PON	RS (248, 216)	RS (248, 216)

NOTE

How to calculate the bandwidth overhead based on the FEC algorithm?

For example, the algorithm RS (248, 232) indicates that the code length is 248 bytes, where 232 bytes are data bytes, 248 – 232 = 16, and 16 bytes are overhead bytes.

Highlight and Application

- Does not require retransmission and provides a high real-time performance.
- Requires an additional bandwidth (Users must balance the transmission quality and bandwidth.).
- Checks and corrects errors after data is transmitted to the peer end, but does not apply to services for which retransmission is enabled.
- Applies to data transmission on the network that has a poor quality.
- Applies to services that have a low requirement on delay (The delay is large if retransmission is configured for services.).

3.5 XG(S)-PON Security Technologies

3.5.1 Line Encryption

Context

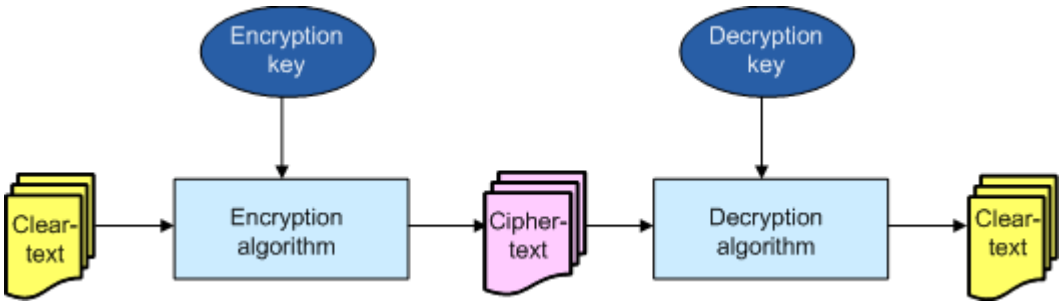
Line encryption technologies are required to eliminate the data theft risk.

The XG(S)-PON supports line encryption in the upstream and downstream directions to ensure two-way secure transmission of data messages.

Working Principle

Figure 3-14 shows line encryption technologies working principle.

Figure 3-14 Line encryption process



 Copyright © Huawei Technologies Co., Ltd.

- The encryption algorithm to be used is the advanced encryption standard (AES).
- The XG(S)-PON systems use the AES-CTR encryption algorithm.

Highlight and Application

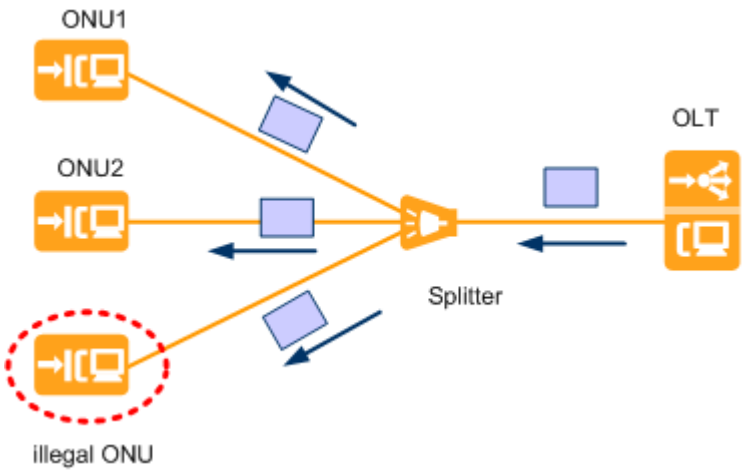
- The line encryption algorithms neither increase overhead nor decrease bandwidth usages.
- The line encryption algorithms will not prolong transmission delays.
- Enable line encryption if the usage scenarios promote high security requirements.

3.5.2 ONU Authentication

Context

In the PON system, downstream data is broadcast to all ONUs. As a result, downstream data destined for certain ONUs or all ONUs may be intercepted by illegal users.

Figure 3-15 Authentication process



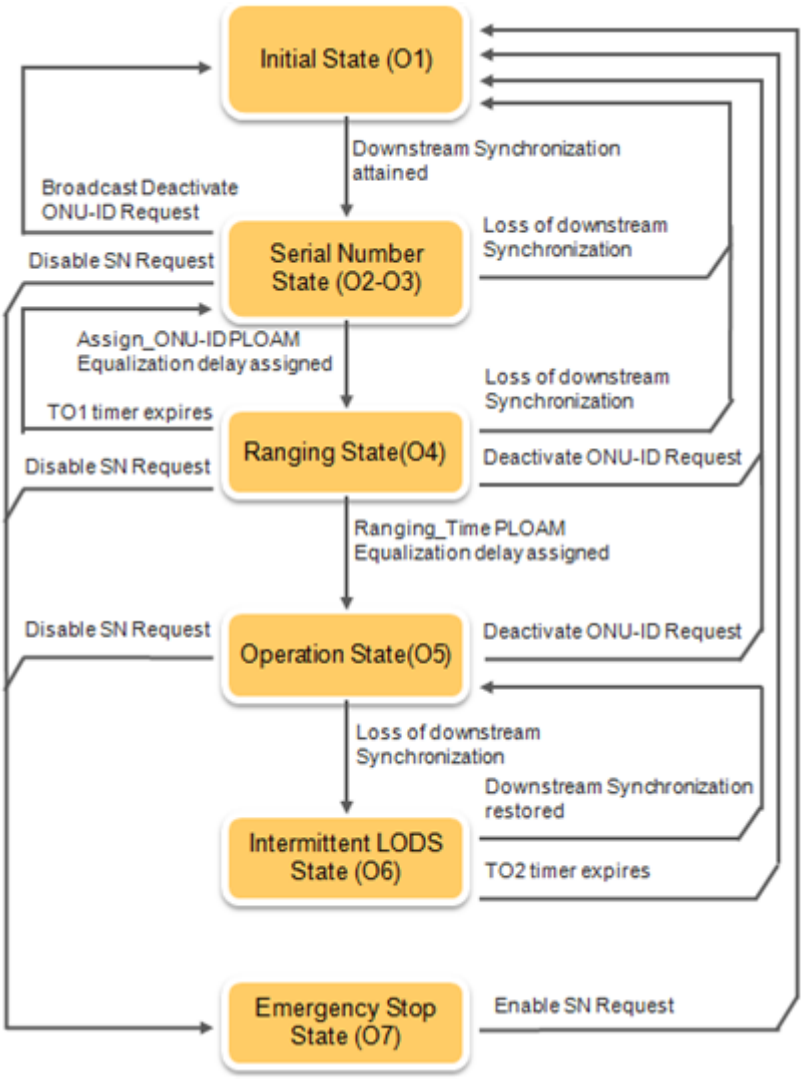
An optical line terminal (OLT) authenticates an optical network unit (ONU) based on the authentication information reported by the ONU and rejects unauthorized ONUs. Only authenticated ONUs can access a passive optical network (PON). This function enables carriers to manage and maintain devices on the PON network flexibly and conveniently.

Authentication Process

Figure 3-16 shows the authentication process:

Copyright © Huawei Technologies Co., Ltd.

Figure 3-16 Authentication Process



The authentication process includes six states:

- O1: Initial state.
- O2-3: Serial Number state.
- O4: Ranging state.
- O5: Operation state.
- O6: Intermittent LODS state.
- O7: Emergency Stop state.

3.5.3 Continuous-Mode ONU Detection

Overview

PON networks use the P2MP (point-to-multipoint) network architecture. They use time division multiple access (TDMA) in the upstream direction. ONUs must send

 Copyright © Huawei Technologies Co., Ltd.

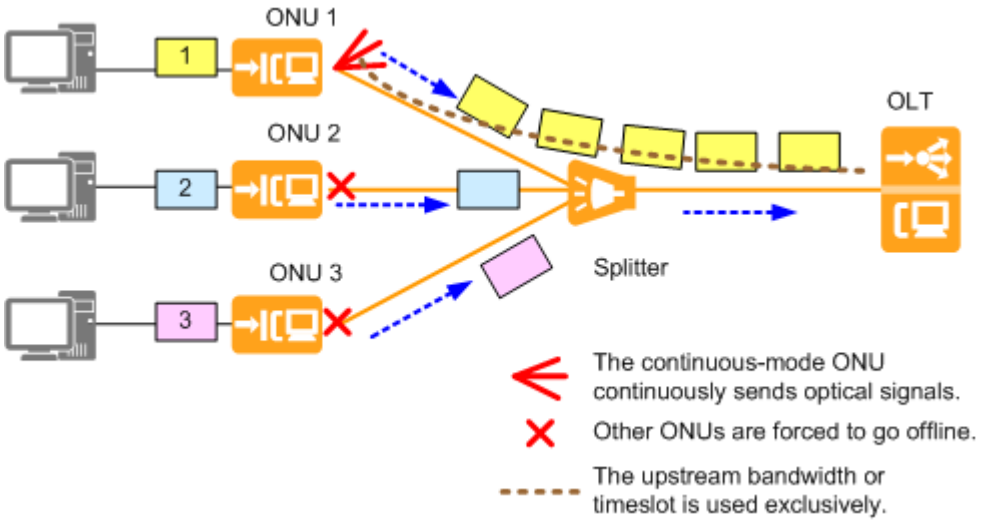
optical signals upstream at the timeslots allocated by the OLT to prevent data conflict.

A continuous-mode ONU is also called a continuous-mode rogue ONU, and is an ONU that transmits optical signals in the upstream direction at any moment.

A continuous-mode ONU adversely affects the system as follows:

- If this ONU has been online, some or all ONUs connected to the same PON port go offline or frequently go offline and online.
- If this ONU has not been configured, other ONUs that have not been configured and connected to the same PON port will not be discovered by the OLT.

Figure 3-17 Continuous-mode ONU



Prerequisites for Rogue ONU detection

- The OLT checks and analyzes the abnormality in the sending of upstream optical signal over a PON line, identifies and isolates rogue ONUs of only non-malicious users. This feature does not apply to the intentionally sabotaged ONU or sub-standard ONU.
- A continuous-mode ONU (rogue ONU) is required to parse and respond to downstream PLOAM messages.
- The OLT can quickly locate a continuous-mode rogue ONU under a PON port only when the ONU supports the standard PLOAM message and its optical module is correctly controlled.

Continuous-Mode ONU Detection

Continuous-mode ONU detection, also called rogue ONU detection, is used for detecting continuous-mode ONUs in the system and isolating them, ensuring proper system running. A continuous-mode ONU detection process involves three stages, checking, detection, and isolation.The three stages are as follows:

 Copyright © Huawei Technologies Co., Ltd.

- Checking: The OLT periodically checks whether a continuous-mode ONU connects to a PON port. This checking cannot locate the continuous-mode ONU.

  The OLT opens an empty gate in the upstream direction to detect ONU optical signals in the upstream direction. If the OLT receives optical signals, it then goes to the detection stage to locate the ONU.

- Detection: The OLT locates the continuous-mode ONU.

  The OLT issues a command to ONUs to instruct the optical modules of the ONUs to send optical signals upstream and checks whether optical signals can be received in the upstream direction. If other ONUs go offline after an ONU sends optical signals, this ONU is a continuous-mode ONU. In a detection process, the OLT checks all ONUs connected to a PON port for detecting all continuous-mode ONUs.

- Isolation: The OLT issues a command to power off the continuous-mode ONU, preventing this ONU from adversely affecting other ONUs connected to the same PON port.

  After an ONU is powered off by the OLT, the ONU cannot send optical signals upstream even after being reset or power recycled. This ONU can send optical signals upstream only after the OLT cancels the isolation.

NOTE

The OLT checks continuous-mode ONUs but does not detect or isolate them by default.

3.6 Evolution From GPON to XG(S)-PON

3.6.1 Network Evolution Overview

For the GPON network that has been deployed on the live network, XG(S)-PON can smoothly evolve from the GPON network to the XG(S)-PON network.

Why Is Smooth Evolution Required?

The smooth evolution from the GPON network to the XG(S)-PON network is mainly due to the following reasons:

- With the development of services, the bandwidth requirement of users is gradually increasing. In the short term, only some users need to be upgraded to the XG(S)-PON network, and some users continue to use the GPON network.
- For a PON network that has been deployed on the live network of a carrier, it is costly to directly cut over the GPON network to the XG(S)-PON network. The evolution from GPON to XG(S)-PON must be completed through smooth evolution.

What are the conditions for smooth evolution from GPON to XG(S)-PON?

XG(S)-PON has the following features to support smooth evolution from a GPON network to a XG(S)-PON network:

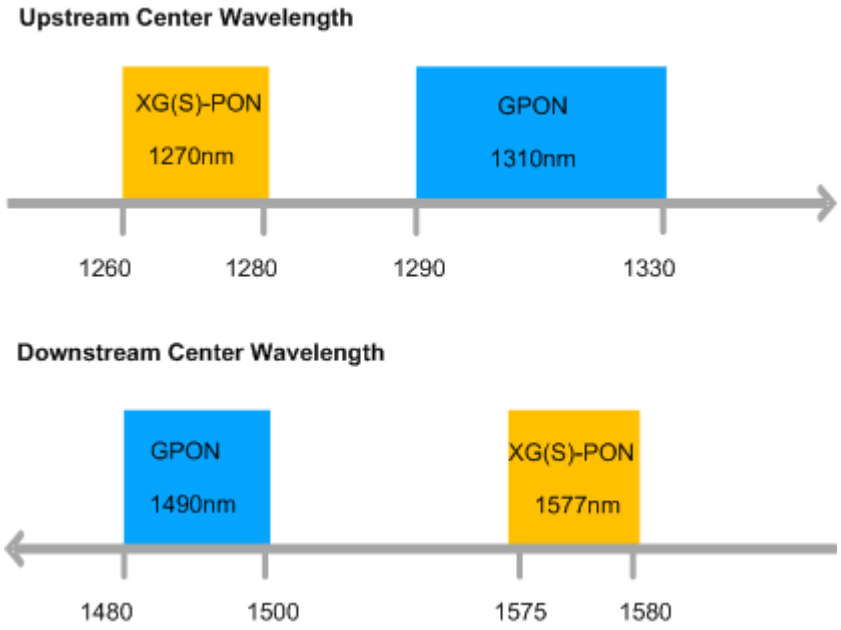
 Copyright © Huawei Technologies Co., Ltd.

- The XG(S)-PON wavelength planning does not overlap the GPON wavelength planning. Therefore, the GPON and XG(S)-PON share the ODN network in the wavelength division multiplexing mode.
- The OLT platform supports coexistence of GPON service boards and XG(S)-PON service boards.
- XG(S)-PON and GPON have the same service provisioning and service deployment operations. They can share the NMS and service provisioning system.

No Wavelength Overlap Between GPON and XG(S)-PON

Figure 3-18 shows the wavelength distribution of GPON and XG(S)-PON.

Figure 3-18 The Wavelength Distribution of GPON and XG(S)-PON

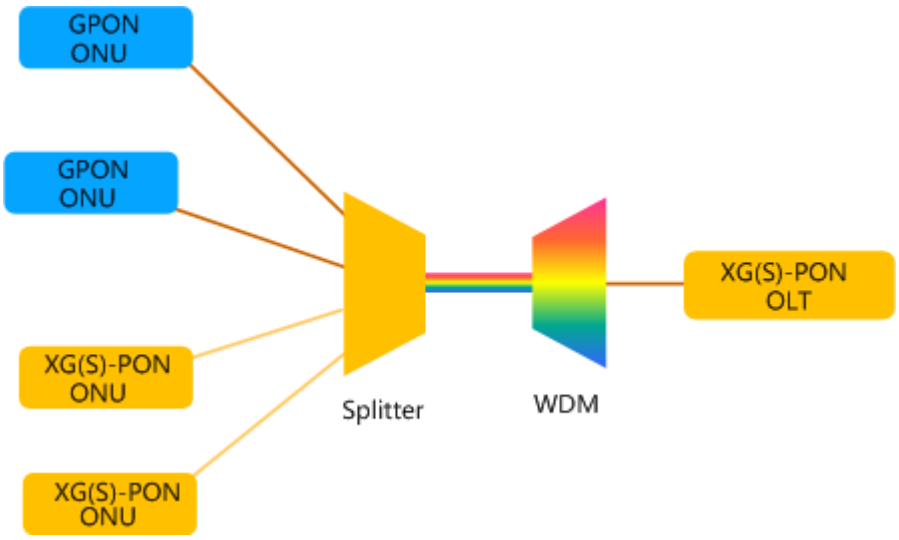


XG(S)-PON Can Co-exists with GPON in the Same ODN

Figure 3-19 shows the GPON and XG(S)-PON can coexist in the same ODN with wavelength division multiplexing (WDM) mode.

 Copyright © Huawei Technologies Co., Ltd.

Figure 3-19 XG(S)-PON Can Co-exists with GPON in the Same ODN



NOTE

WDM: Wavelength-Division Multiplexing

3.6.2 Evolution Solution - Using XG(S)-PON Combo

This solution introduction smooth evolution from the GPON network to the XG(S)-PON network by adding the XG(S)-PON Combo Service Board.

3.6.2.1 What Is XG(S)-PON Combo?

The PON Combo supports two PON technologies through a combo port. One PON Combo port supports multiple types of ONUs at the same time, implementing hybrid network construction of GPON and XG(S)-PON and smooth evolution from GPON to XG(S)-PON.

Huawei PON Combo provides two types: XG-PON Combo and XGS-PON Combo.

- XG-PON Combo:
  - XG-PON Combo means that GPON and XG-PON are integrated into one port through Combo.
  - The XG-PON Combo supports the access of GPON ONUs and XG-PON ONU.
- XGS-PON Combo:
  - XGS-PON Combo means that GPON and XGS-PON are integrated into one port through Combo.
  - The XGS-PON Combo supports GPON ONU, XG-PON ONU, and XGS-PON ONU.

  NOTE

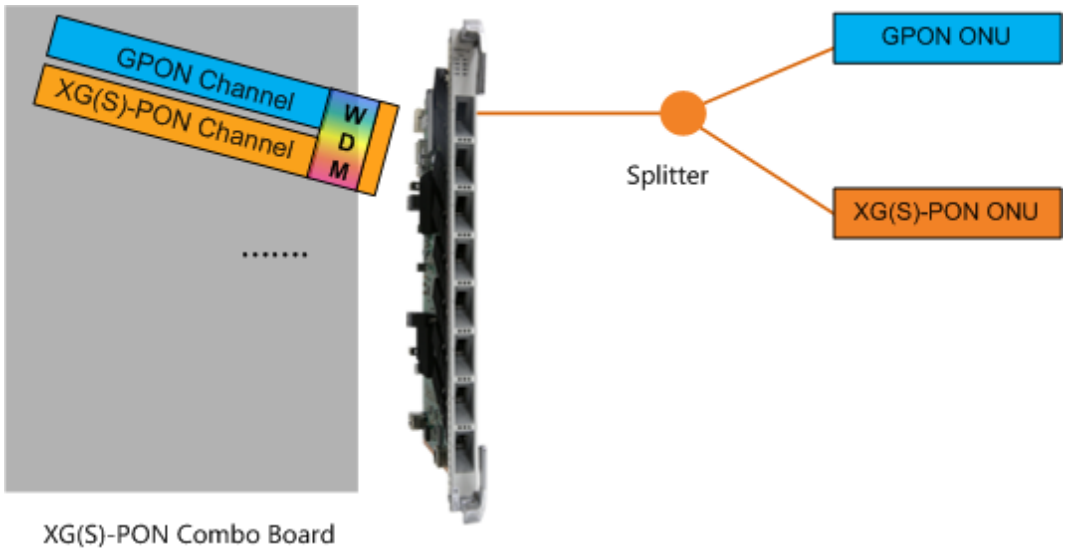
  In this document, XG-PON Combo and XGS-PON Combo are short for XG(S)-PON Combo.

 Copyright © Huawei Technologies Co., Ltd.

3.6.2.2 XG(S)-PON Combo Working Principle

Figure 3-20 shows the working principle of the XG(S)-PON combo service board.

Figure 3-20 Working Principle of the XG(S)-PON Combo Service Board



Working Principle:

- The XG(S)-PON combo board works with the XG(S)-PON combo optical module to implement a combo port working in GPON mode and XG(S)-PON mode at the same time.
- The XG(S)-PON combo optical module integrates the GPON optical module, XG(S)-PON optical module, and WDM multiplexer to share ODN resources without adding an external WDM multiplexer.
  - In the upstream direction: After the upstream signal of the ONU enters the XG(S)-PON combo port, the WDM determines whether the signal is a GPON signal or a XG(S)-PON signal according to the wavelength, and then sends the signal to the corresponding channel for processing.
  - In the downstream direction: After the downstream signal is sent from the OLT XG(S)-PON combo port, the GPON ONU and the XG(S)-PON ONU select the wavelength through the internal wavelength filter to receive the signal.

3.6.2.3 XG(S)-PON Combo Evolution Solution

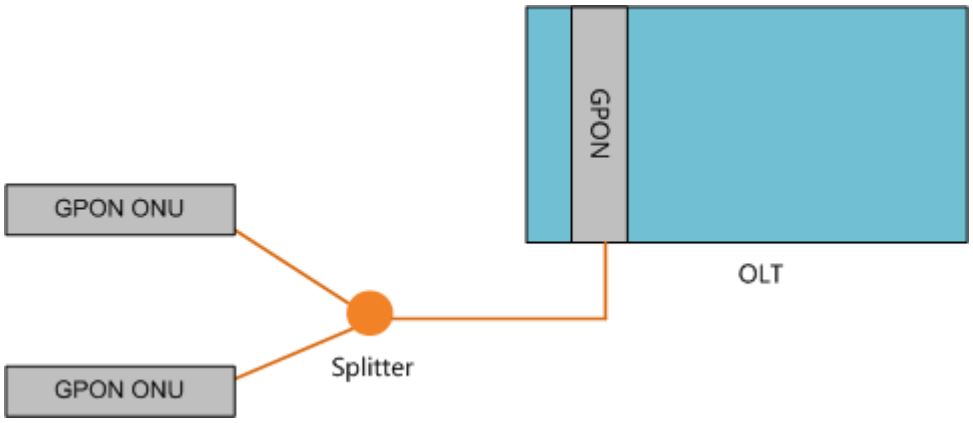
This solution introduction smooth evolution from the GPON network to the XG(S)-PON network by adding the XG(S)-PON Combo Service Board.

Before Evolution

Figure1 shows the GPON FTTx network, the OLT uses the GPON service board to connect to the GPON ONU before the evolution.

Copyright © Huawei Technologies Co., Ltd.

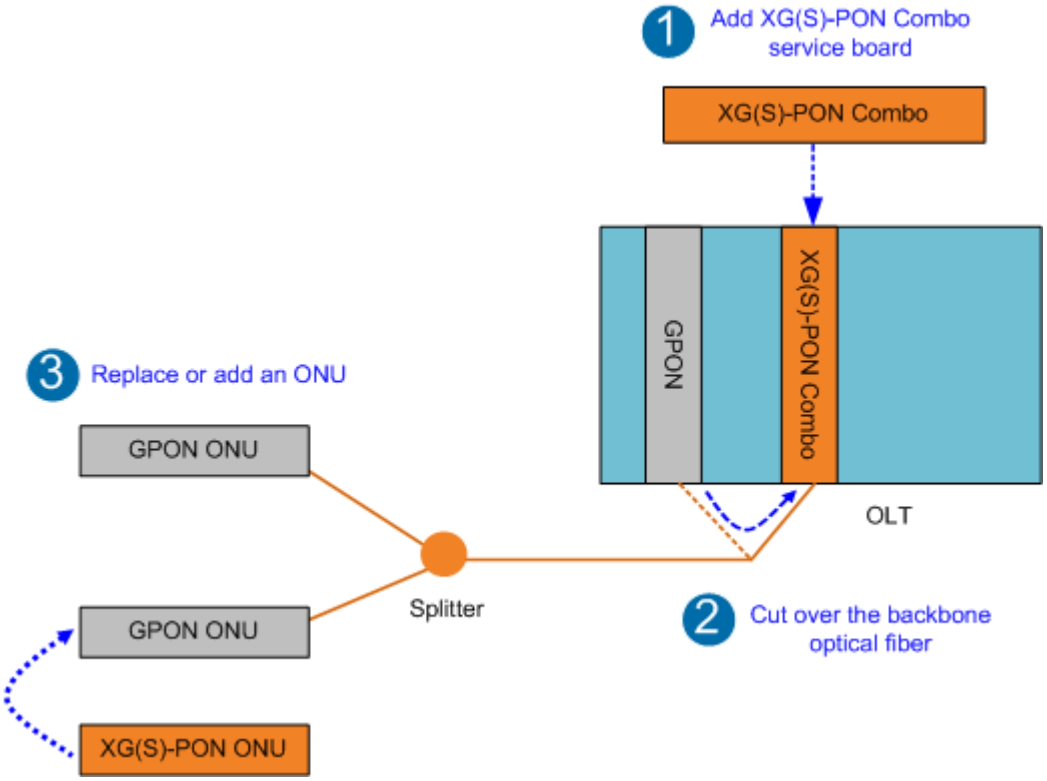
Figure 3-21 GPON FTTx Network



Evolution Solution

Figure2 shows the smooth evolution from the GPON network to the XG(S)-PON network by adding the XG(S)-PON Combo Service Board.

Figure 3-22 Evolution Solution



NOTE

XG(S)-PON is short for XG-PON and XGS-PON.

Operation Procerure

1. Add XG(S)-PON Combo service board.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

- If the OLT slot on the live network supports only 10 Gbit/s switching capacity and cannot meet the forwarding requirements of XG(S)-PON service boards, replace the subrack.
- If the upstream port of the OLT supports only GE upstream bandwidth, it is recommended that you replace the upstream port with 10GE upstream bandwidth or even larger bandwidth to prevent the upstream port from becoming a bottleneck.

2. Migrate the backbone optical fiber of the GPON port to the XG(S)-PON combo port.
3. Add a XG(S)-PON ONU or replace a GPON ONU.

NOTE

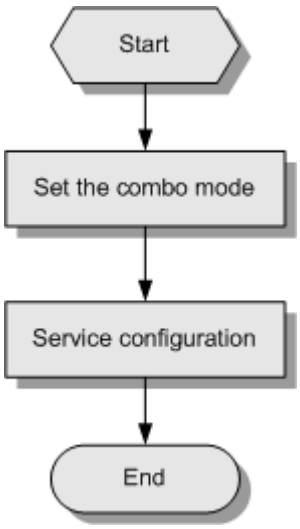
- The XG-PON Combo optical module can be connected to GPON ONU and XG-PON ONU.
- The XGS-PON Combo optical module can be connected to GPON ONU, XG-PON ONU and XGS-PON ONU.
- The ranging compensation distance (GPON ONU and XG(S)-PON ONU) planned for the XG(S)-PON Combo port must be the same.

4. After the GPON ONU is replaced with the XG(S)-PON ONU, the evolution is complete.

3.6.2.4 XG(S)-PON Combo Configuration Guide

Figure 1 shows the process of configuring the XG(S)-PON Combo service. After the XG(S)-PON Combo mode is configured, you can configure the service according to the configuration cases in the corresponding scenario.

Figure 3-23 XG(S)-PON Combo Service Configuration Process



Configuration Procedure

1. Set the combo mode of a XG(S)-PON Combo port.

   On the OLT, run the **port combo-mode** { **all** | *portid* } { **auto** | **wdm** } command to set the combo mode of a XG(S)-PON Combo port.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

- **auto**: sets the combo mode of a XG(S)-PON Combo PON port to **auto**.
- **wdm**: sets the combo mode of a XG(S)-PON Combo PON port to **wdm**.
- The XG(S)-PON Combo port mode must be the same as that of the XG(S)-PON Combo optical module.

For example, set the combo port 0/1/0 to the auto mode.

```
huawei(config-if-gpon-0/3)#port combo-mode
{ all<K>|portid<U><0,15> }:0
{ auto<K>|wdm<K> }:auto
```

2. Service Configuration. For details, refer to the <Configuration Guide>.

NOTE

After the GPON is switched to the XG(S)-PON, pay attention to the following points:

- If the GPON ONU is not replaced after the evolution, the GPON ONU configuration remains unchanged.
- If GPON ONU is replaced with XG-PON ONU or XGS-PON ONU after the evolution, delete GPON ONU and configure XG-PON ONU or XGS-PON ONU.

– For an XG(S)-PON FTTH network, configure services according to the XG(S)-PON FTTH Configuration Guide.
  - For carrier customers, visit: **XG(S)-PON FTTH Configuration Guide**
  - For enterprise customers, visit: **XG(S)-PON FTTH Configuration Guide**
– For the XG(S)-PON FTTB/C networking, configure services according to the XG(S)-PON FTTB/C Configuration Guide.
  - For carrier customers, visit: **XG(S)-PON FTTB/C Configuration Guide**
  - For enterprise customers, visit: **XG(S)-PON FTTB/C Configuration Guide**

3.6.3 Evolution Solution-Using Flex-PON

This solution describes the smooth evolution from the GPON network to the XG(S)-PON network using Flex-PON service boards.

3.6.3.1 What Is Flex-PON?

Flex-PON is a flexible PON solution developed by Huawei for GPON and XG(S)-PON. It implements smooth network evolution through the integrated Flex-PON service board.

Multi-mode Integration

The ports of the Flex-PON service board can work in multiple modes, including:

- GPON
- XG-PON

 Copyright © Huawei Technologies Co., Ltd.

- XGS-PON
- XG-PON/GPON Combo
- XGS-PON/GPON Combo

After the OLT obtains the corresponding RTU authorization, the Flex-PON port can work in the corresponding mode.

Smooth Evolution

Flex-PON supports the evolution from a GPON network to an XG(S)-PON network.

- You only need to load the right to use (RTU) and replace the optical module to realize the evolution from GPON to XG(S)-PON without replacing the board.
- The OSS needs to be integrated only once. After GPON evolves to XG(S)-PON, the OSS does not need to be integrated again.
- Replace ONUs as required.

3.6.3.2 What Is 10G PON Port RTU?

Right to use (RTU) is a license that controls the working mode of a Flex-PON port. The working mode of a Flex-PON port can be changed by using different RTUs.

Mapping between the Flex-PON port mode, RTU type, Optical module type, and ONU type

When formulating an evolution solution for the Flex-PON, ensure that the port mode is supported by the RTU, optical module, and ONU. The following table lists the mapping between the Flex-PON port mode and the RTU, optical module, and ONU type.

Table 3-2 Mapping Between the Port Mode and the RTU, Optical Module, and ONU

Port Mode	RTU Type	Optical Module Type	ONU Type
GPON	No need RTU	GPON	GPON ONU
XG-PON	10G PON RTU	XG-PON	XG-PON ONU
XGS-PON	10G PON RTU	XGS-PON	• XG-PON ONU • XGS-PON ONU
XG-PON/ GPON Combo	10G PON RTU	XG-PON/GPON Combo	• GPON ONU • XG-PON ONU
XG-PON/ GPON Combo	No RTU	XG-PON/GPON Combo	GPON ONU

 Copyright © Huawei Technologies Co., Ltd.

Port Mode	RTU Type	Optical Module Type	ONU Type
XGS-PON/GPON Combo	10G PON RTU	XGS-PON/GPON Combo	• GPON ONU • XG-PON ONU • XGS-PON ONU
XGS-PON/GPON Combo	No RTU	XGS-PON/GPON Combo	GPON ONU

RTU Authorization Process

The process for NCE to authorize RTUs is as follows:

1. NCE polls the total RTU consumption of all managed OLTs at a specified time every day.
2. NCE notifies all OLTs of whether RTUs are sufficient based on the polling result:
   - If the RTUs are sufficient, a message indicating that the RTUs are sufficient is sent to all OLTs.
   - If the RTUs are overloaded (exhausted), the RTU overload information is sent to all OLTs.
3. The OLT triggers RTU authorization:

   Run the **port flex-mode** command on the OLT to configure the Flex-PON port mode and insert an optical module.
4. After the OLT triggers RTU authorization, the OLT determines whether to authorize the RTU based on the notification received from NCE (in step 2, NCE sends the RTU polling result to the OLT):
   - If the RTU is sufficient, the RTU is granted to the Flex-PON port that requests authorization. In this case, the Flex-PON port can work normally in the target mode.
   - If the RTU is overloaded (exhausted), the RTU is not granted to the Flex-PON port that requests authorization. In this case, the Flex-PON port cannot work in the target mode.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

You can run the display port state command on the OLT to check whether the RTU is authorized successfully and check the value of **Port RTU**.

Example:

To query the status information about a port 0/3/0, do as follows:

```
huawei(config-if-gpon-0/3)#display port state 0

---------------------------------------------------------------------------
  Port Information
---------------------------------------------------------------------------
  F/S/P                     0/3/0
  Port state                Offline
  Last down cause           -
  Last up time              -
  Last down time            -
  Port energy-saving flag   No
  xPON MAC chipset state    Normal
  Signal detect             Normal
  Available bandwidth(Kbps) 1241088
  Illegal rogue ONT         Inexistent
  OTDR test status          -
  Upstream bandwith RTU     -
  Downstream bandwith RTU   -
  Port RTU                  GPON        //Indicates that the RTU of the port is GPON.
```

FAQ

Describes the FAQs related to 10G PON RTU.

Table 3-3 The FAQs related to 10G PON RTU.

Question	Answer
What are the RTU types?	10G PON RTU on the Flex-PON board: • Part number: 88035APW • Description: 10G PON Port RTU, Based on per PON Port.
Are RTUs managed at the port level?	Yes.
What are the symptoms of a Flex-PON port without an RTU enabled?	If the Flex-PON port of the RTU is disabled, the ONU cannot go online. • The port works in GPON mode. When the optical module is XG(S)-PON/GPON Combo, only GPON ONUs can go online. XG(S)-PON ONUs are not allowed to go online. • When the Flex-PON port works in XG(S)-PON/GPON mode and the optical module is XG(S)-PON/GPON Combo, only GPON ONUs can go online. XG(S)-PON ONUs are not allowed to go online.

 Copyright © Huawei Technologies Co., Ltd.

Question	Answer
Will services be interrupted when the working mode of a Flex-PON port is switched?	• When a GPON optical module is deployed on a Flex-PON port, switch the Flex-PON port mode from GPON to XG(S)-PON/GPON Combo mode. The GPON ONU can go online regardless of whether RTUs are configured. • When an XG(S)-PON/GPON Combo optical module is deployed on a Flex-PON port and the Flex-PON port mode is switched from GPON to XG(S)-PON/GPON Combo mode, the following situations occur: – If there is no RTU, the XG(S)-PON ONU cannot go online, but the GPON ONU goes online normally. – If there are RTUs, the XG(S)-PON ONU and GPON ONU can go online normally.
After RTUs are activated for all ports on a Flex-PON board of an OLT, are RTUs unaware of the board migration if the board is migrated to another OLT?	RTUs are a part of Flex-PON boards. In the same NCE domain, RTUs can be migrated without being perceived by hardware boards. (After board migration, the number of RTUs consumed by the original OLT is deducted, and the number of RTUs consumed by the new OLT is increased.)
Is an RTU unaware of board replacement after a faulty Flex-PON board on an OLT is replaced?	• If the RTUs are overdrawn and the grace period expires, the customer needs to apply for a new RTU after replacing the board. • If the RTUs are not used up or is overdrawn but within the grace period, the customer does not need to apply for an RTU after replacing the board. Instead, the original RTU license is used.
On an OLT, after an XGS-PON optical module is migrated from Flex-PON board 1 to Flex-PON board 2, is the RTU unaware of the migration?	The RTU is a part of the Flex-PON board but not a part of the optical module. If only the optical module is migrated, the RTU is not migrated.
After a faulty OLT is replaced and the original OLT database is loaded, is the RTU unaware of the replacement?	RTUs are migrated with boards but not with OLTs. RTUs are unaware of OLT replacement.

 Copyright © Huawei Technologies Co., Ltd.

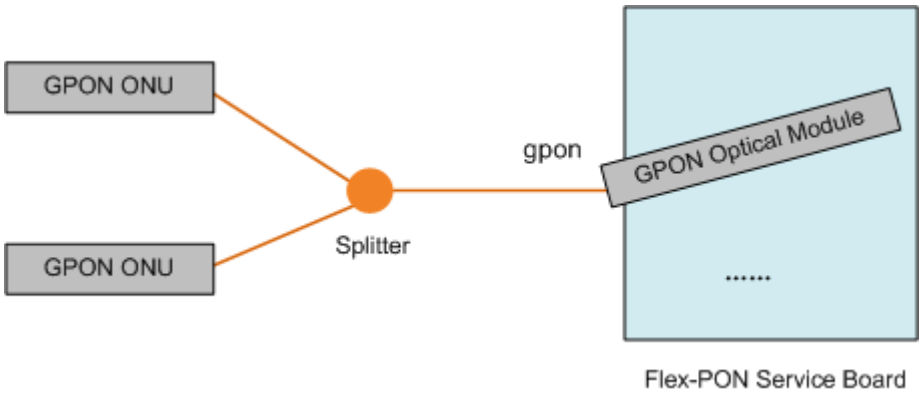
Question	Answer
If an OLT or Flex-PON board of a site is removed from the network, can the RTU be reclaimed to the NCE RTU resource pool?	• After a board exits from the network, the RTU used by the board is reclaimed. • After an OLT is removed from the network (that is, the NE is deleted from NCE), the RTU used by the OLT is reclaimed. If a board on the OLT is used on another OLT, the number of used RTUs is re-calculated.
After a Flex-PON port on an OLT is disabled and the board is removed, will the RTU be reclaimed to the RTU resource pool?	After a Flex-PON board is removed, the RTUs used by the board are reclaimed. After the board is re-inserted, the number of used RTUs is re-calculated.
Will RTU activation and management be affected when NCE is offline?	The status is the same as that before NCE goes offline.
Can an XGS-PON ONU Go Online Without an RTU?	If there is no 10G PON port RTU, the XGS-PON ONU is forced to go offline after going online, and the ONU goes offline because the port lacks RTU is reported.

3.6.3.3 Evolution Solution Overview

Before Evolution

Figure 1 shows the GPON FTTx network, the OLT uses the Flex-PON service board to connect to the GPON ONU before the evolution.

Figure 3-24 GPON FTTx Network

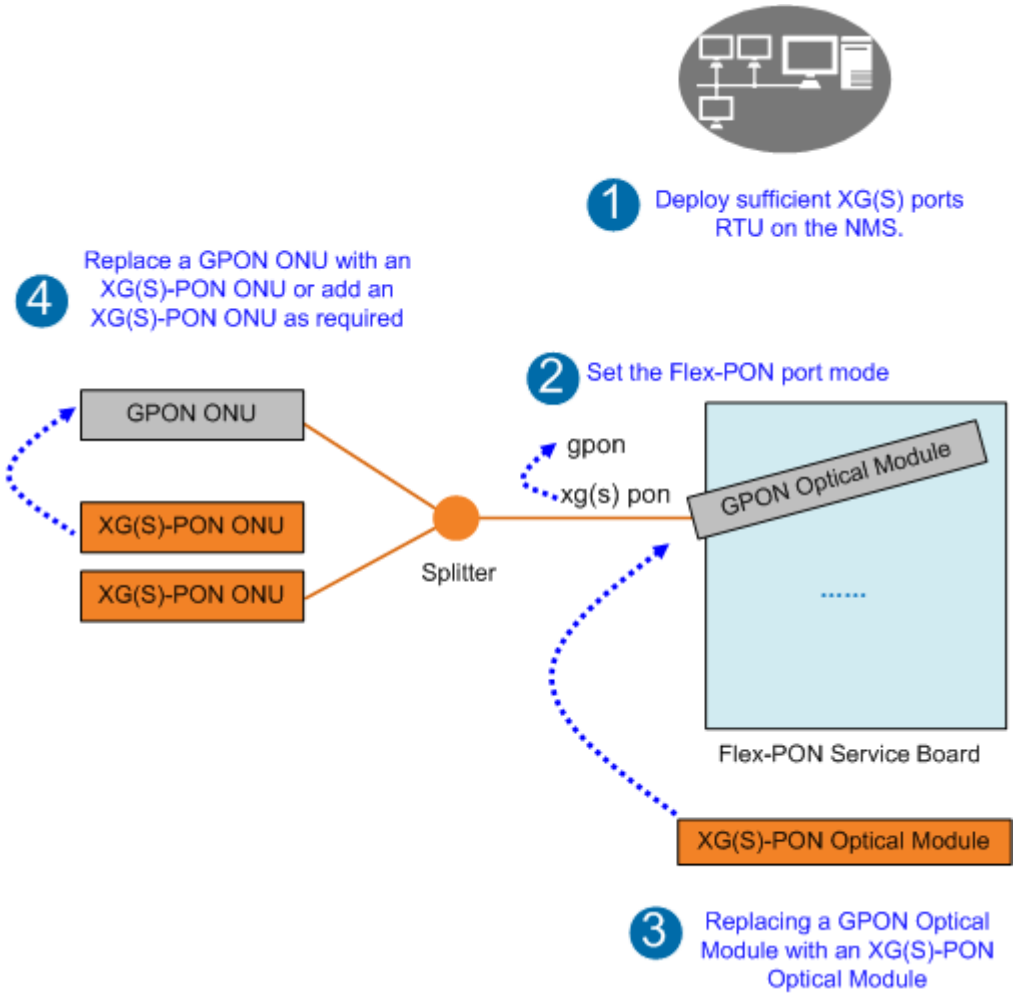


Evolution Solution

Figure 2 shows the smooth evolution from the GPON network to the XG(S)-PON network by Flex-PON solution.

 Copyright © Huawei Technologies Co., Ltd.

Figure 3-25 Evolution Solution



The evolution solution must be completed in five steps:

1. Deploy sufficient XG(S) ports RTU on the NCE;
2. Configure the port mode. (This step triggers RTU authorization by the NCE.);
3. Authorized RTU by NCE (This step does not require manual operations. If the RTUs on the U2000 are sufficient, NCE grants the RTUs to the Flex-PON port that requests authorization. After the RTUs are granted, the port can switch to the target mode. If the RTUs on NCE are overloaded or exhausted, , the RTU is not granted to the port. In this case, the port cannot switch to the target mode.
4. Replace the optical module as required;
5. Replace or add ONUs as required;
6. Configuring Services;

Operation Procedure

1. Deploy sufficient XG(S) ports RTU on the NCE.

 Copyright © Huawei Technologies Co., Ltd.

NOTICE

- Deploying a sufficient number of RTUs on XG(S) ports on NCE is a prerequisite for successful Flex-PON port mode switching.
- The RTU can be loaded only through the NMS, this service scenario cannot be performed without the NMS.

2. Set the Flex-PON port mode. On the OLT, run the **port flex-mode** { **all** | *portid* } { **gpon** | **xg-pon** | **xgs-pon** | **combo-pon** { **xg/g** | **xgs/g** } [**auto** | **wdm**] }command to set the port mode of the Flex-PON.

Parameter	Parameter Description
all	Sets the port mode of all ports on a flex-PON board.
portid	Set the port mode of a Flex-PON port.
gpon	Set the port mode of the Flex-PON port to **GPON**.
xg-pon	Set the port mode of the Flex-PON port to **XG-PON**.
xgs-pon	Set the port mode of the Flex-PON port to **XGS-PON**.
combo-pon	Set the port mode of the Flex-PON port to **Combo-PON**. In this case, you need to specify the Combo-PON type.
xg/g	Set the port mode of the Flex-PON port to **XG-PON&GPON Combo-PON**.
xgs/g	Set the port mode of the Flex-PON port to **XGS-PON&GPON Combo-PON**.
auto	Set the multiplexing mode of a combo-PON port to **auto**.
wdm	Set the multiplexing mode of a combo-PON port to **WDM.**

For example, set port 0 of the Flex-PON board to work in xgs/g combo mode and set the multiplexing mode to WDM.

```
huawei(config-if-gpon-0/1)#port flex-mode
{ all<K>|portid<U><0,15> }:0
{ combo-pon<K>|gpon<K>|xg-pon<K>|xgs-pon<K> }:combo-pon
{ xg/g<K>|xgs/g<K> }:xgs/g
{ <cr>|auto<K>|wdm<K> }:wdm
```

3. Replace the GPON optical module with a XG(S)-PON optical module.

NOTICE

The optical module type must be the same as the port mode. Otherwise, the optical module cannot work properly.

 Copyright © Huawei Technologies Co., Ltd.

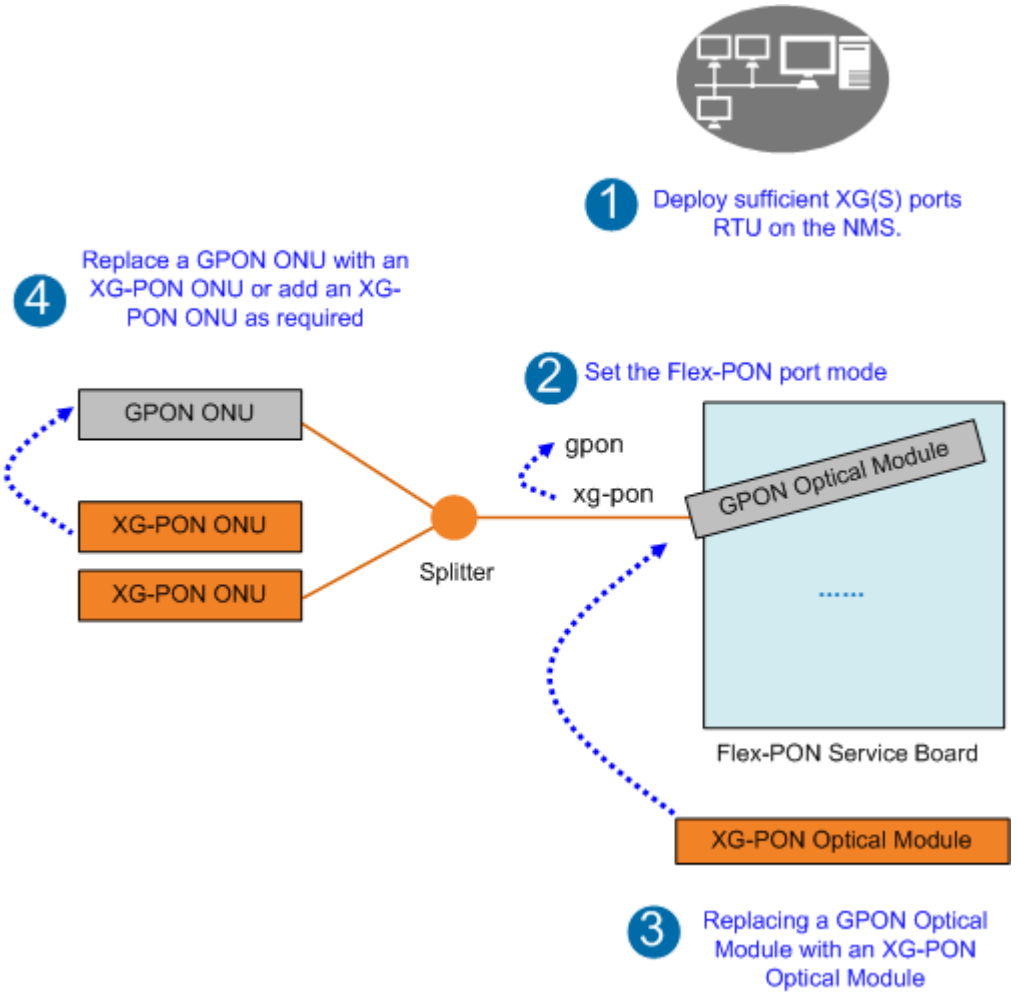
4. Add the XG(S)-PON ONU or replace the GPON ONU with the XG(S)-PON ONU.
5. Configure services. For details, see the **Flex-PON Configuration Guide**.

3.6.3.4 GPON Evolution to XG-PON

Evolution Solution

Figure 1 shows the smooth evolution from the GPON network to the XG-PON network by Flex-PON solution.

Figure 3-26 GPON Evolution to XG-PON



Operation Procedure

1. Deploy sufficient XG(S) ports RTU on the NCE.
2. Set the Flex-PON port mode. On the OLT, set the Flex-PON port mode from **gpon** to **xg-pon**.

   For example, set port 0 of the Flex-PON board to the xg-pon mode.

```
huawei(config-if-gpon-0/1)#port flex-mode
{ all<K>|portid<U><0,15> }:0
{ combo-pon<K>|gpon<K>|xg-pon<K>|xgs-pon<K> }:xg-pon
```

 Copyright © Huawei Technologies Co., Ltd.

NOTICE

If a GPON ONU already exists on the Flex-PON port, you need to delete the GPON ONU and then switch the mode. Otherwise, the switching fails.

3. Replace the GPON optical module with a XG-PON optical module.
4. Add the XG-PON ONU or replace the GPON ONU with the XG-PON ONU.

   NOTE

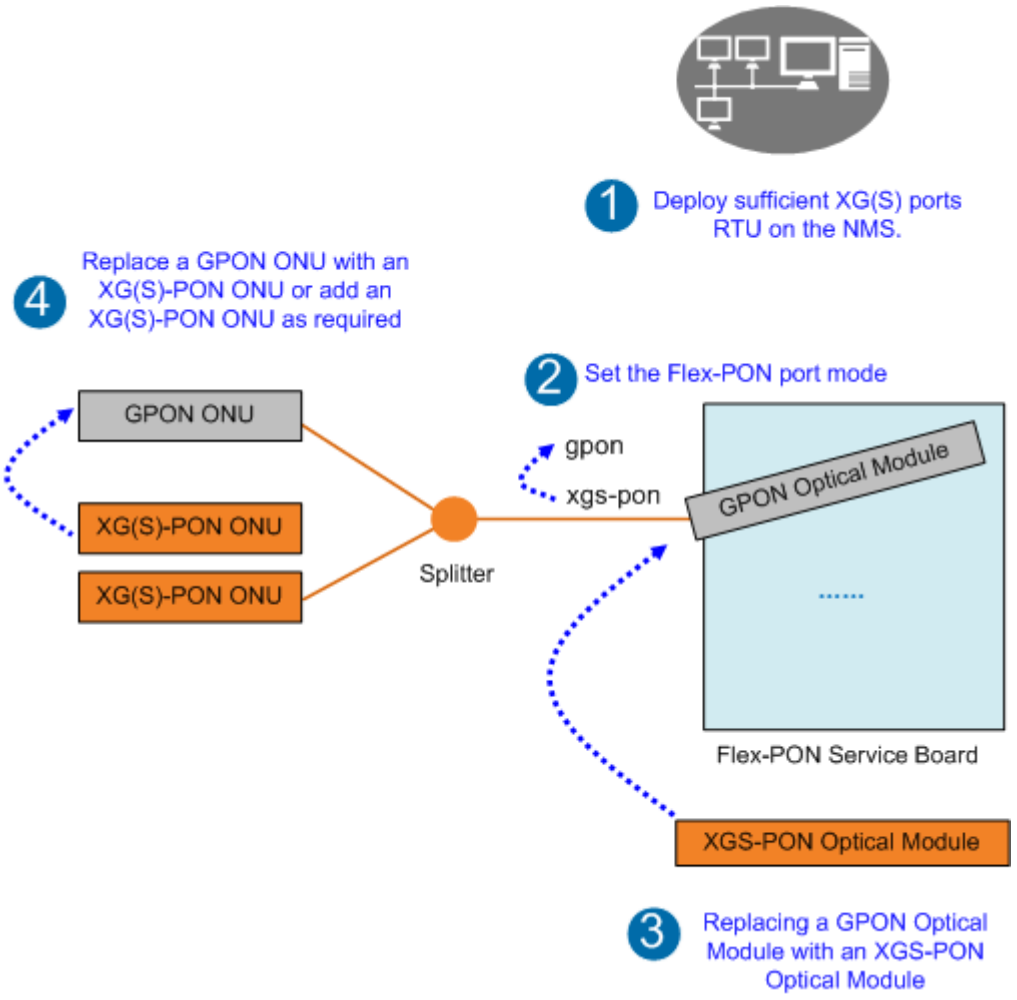
   The XG-PON optical module can be connected to only the XG-PON ONU.
5. Configure services. For details, see the **Flex-PON Configuration Guide**.
6. The evolution is complete.

3.6.3.5 GPON Evolution to XGS-PON

Evolution Solution

Figure 1 shows the smooth evolution from the GPON network to the XGS-PON network by Flex-PON solution.

Figure 3-27 GPON Evolution to XGS-PON



 Copyright © Huawei Technologies Co., Ltd.

Operation Procedure

1. Deploy sufficient XG(S) ports RTU on the NCE.
2. Set the Flex-PON port mode. On the OLT, set the Flex-PON port mode from **gpon** to **xgs-pon**.

   For example, set port 0 of the Flex-PON board to the xgs-pon mode.

   ```
   huawei(config-if-gpon-0/1)#port flex-mode
   { all<K>|portid<U><0,15> }:0
   { combo-pon<K>|gpon<K>|xg-pon<K>|xgs-pon<K> }:xgs-pon
   ```

   NOTICE

   If a GPON ONU already exists on the Flex-PON port, you need to delete the GPON ONU and then switch the mode. Otherwise, the switching fails.
3. Replace the GPON optical module with a XGS-PON optical module.
4. Add the XGS-PON ONU or replace the GPON ONU with the XGS-PON ONU.

   NOTE

   The XGS-PON optical module can be connected to XG-PON ONU and XGS-PON ONU.
5. Configure services. For details, see the **Flex-PON Configuration Guide**.
6. The evolution is complete.

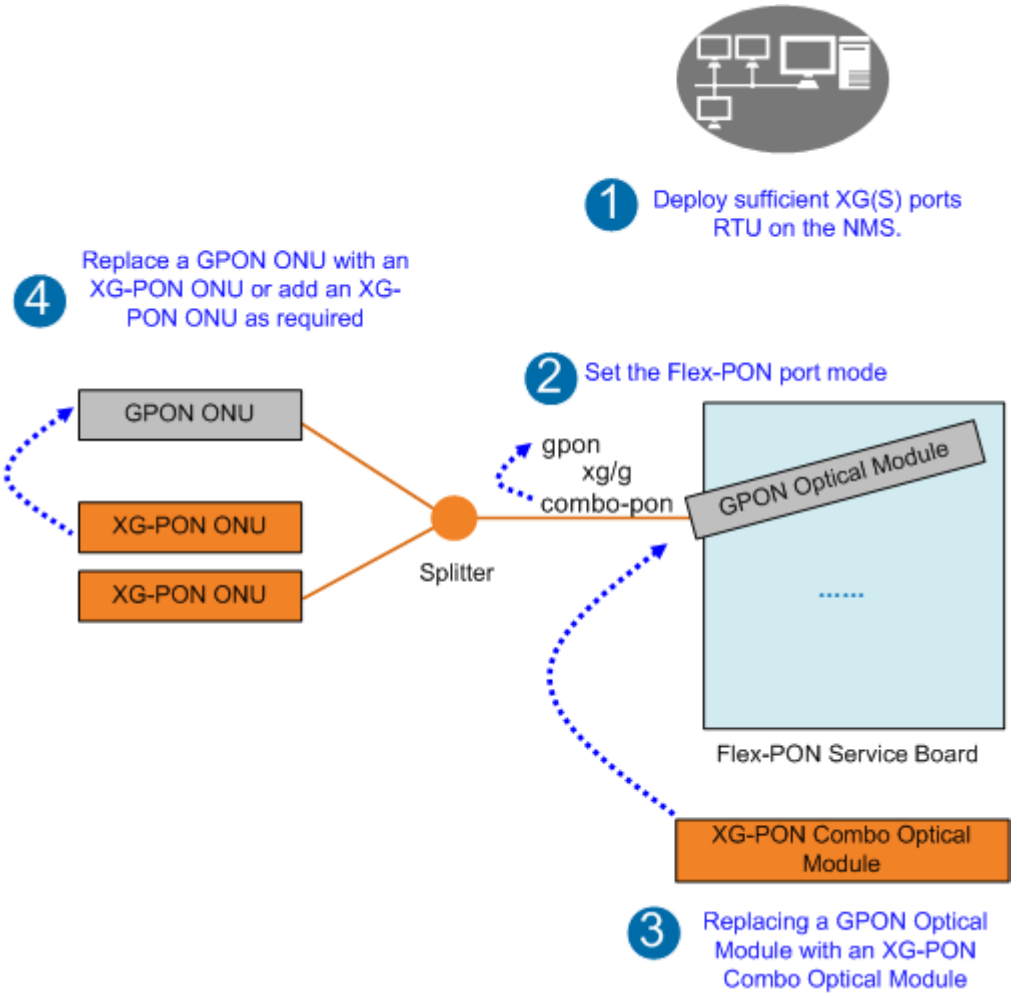
3.6.3.6 GPON Evolution to XG-PON Combo

Evolution Solution

Figure 1 shows the smooth evolution from the GPON network to the XG-PON Combo network by Flex-PON solution.

 Copyright © Huawei Technologies Co., Ltd.

Figure 3-28 GPON Evolution to XG-PON Combo



Operation Procedure

1. Deploy sufficient XG(S) ports RTU on the NCE.
2. Set the Flex-PON port mode. On the OLT, set the Flex-PON port mode from **gpon** to **xg/g combo-pon** and set the multiplexing mode to WDM.

```
huawei(config-if-gpon-0/1)#port flex-mode
{ all<K>|portid<U><0,15> }:0
{ combo-pon<K>|gpon<K>|xg-pon<K>|xgs-pon<K> }:combo-pon
{ xg/g<K>|xgs/g<K> }:xg/g
{ <cr>|auto<K>|wdm<K> }:wdm
```

   NOTE

   - Before the switchover, the ONU configured in GPON port mode is set to Auto. After the switchover, both GPON ONUs and XG-PON ONUs can go online through authentication.
   - To avoid additional bandwidth consumption, you are advised to change ont-type to 10g/2.5g (XG-PON ONU) or 2.5g/1.25g (GPON ONU) based on the actual ONU type.

3. Replace the GPON optical module with a XG-PON Combo optical module.

 Copyright © Huawei Technologies Co., Ltd.

NOTICE

- After the port mode is switched to GPON&XG-PON Combo mode, the GPON optical module can still work and the GPON ONU can work normally.
- When the port mode is GPON, replace the optical module with a GPON&XG-PON combo optical module. The GPON ONU can work normally.
- When the port mode is GPON&XG-PON combo mode and the optical module is GPON&XG-PON combo optical module, the port can work normally only when there are XG(S) ports RTU.

4. Add the XG-PON ONU or replace the GPON ONU with the XG-PON ONU.

   NOTE

   The XG-PON Combo optical module can be connected to GPON ONU and XG-PON ONU.

5. Configure services. For details, see the **Flex-PON Configuration Guide**.
6. The evolution is complete.

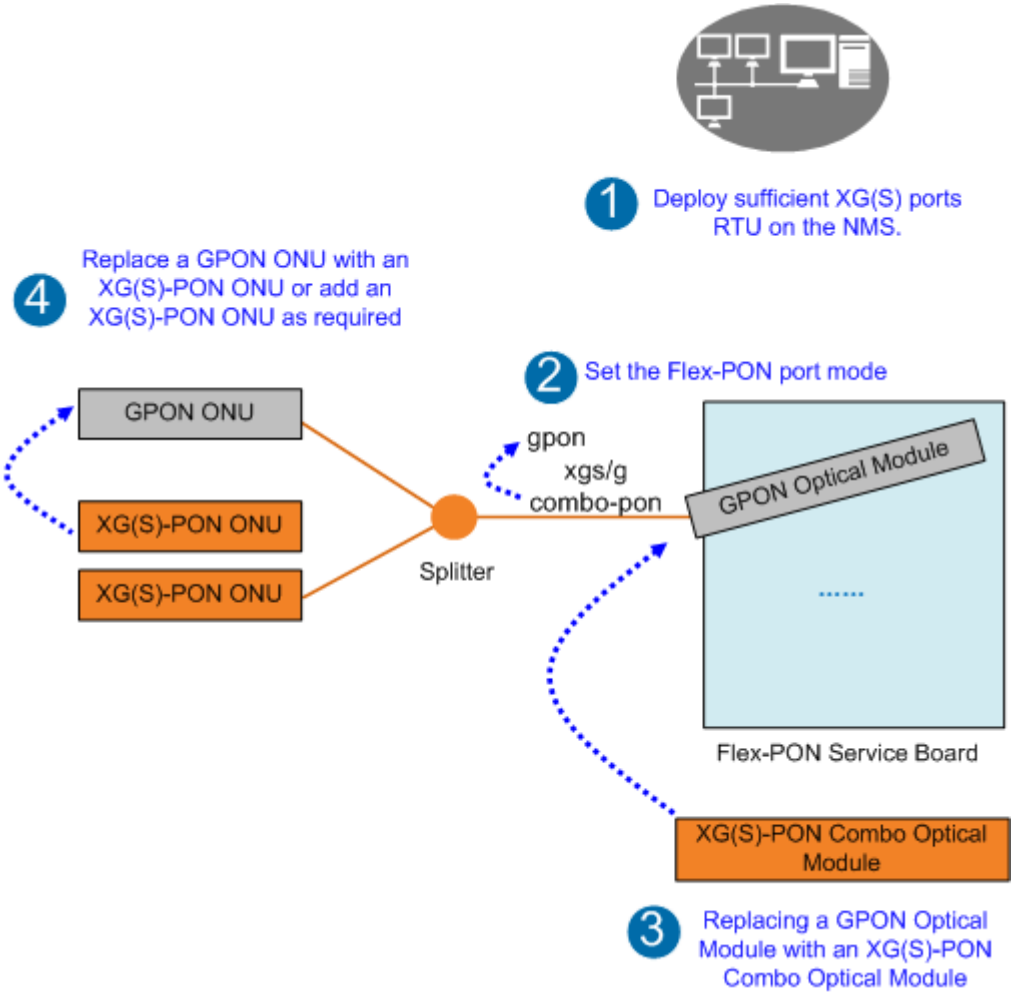
3.6.3.7 GPON Evolution to XGS-PON Combo

Evolution Solution

Figure 1 shows the smooth evolution from the GPON network to the XGS-PON Combo network by Flex-PON solution.

Copyright © Huawei Technologies Co., Ltd.

Figure 3-29 GPON Evolution to XGS-PON Combo



NOTE

XG(S)-PON is short for XG-PON and XGS-PON.

Operation Procedure

1. Deploy sufficient XG(S) ports RTU on the NCE.
2. Set the Flex-PON port mode. On the OLT, set the Flex-PON port mode from **gpon** to **xgs/g combo-pon** and set the multiplexing mode to WDM.

```
huawei(config-if-gpon-0/1)#port flex-mode
{ all<K>|portid<U><0,15> }:0
{ combo-pon<K>|gpon<K>|xg-pon<K>|xgs-pon<K> }:combo-pon
{ xg/g<K>|xgs/g<K> }:xgs/g
{ <cr>|auto<K>|wdm<K> }:wdm
```

NOTE

- Before the switchover, the ONU configured in GPON port mode is set to Auto. After the switchover, both GPON ONUs and XG(S)-PON ONUs can go online through authentication.
- To avoid additional bandwidth consumption, you are advised to change ont-type to 10g/2.5g (XG-PON ONU), 10g/10g (XGS-PON ONU), or 2.5g/1.25g (GPON ONU) based on the actual ONU type.

 Copyright © Huawei Technologies Co., Ltd.

3. Replace the GPON optical module with a XGS-PON Combo optical module.

NOTICE

- After the port mode is switched to GPON&XGS-PON Combo mode, the GPON optical module can still work and the GPON ONU can work normally.
- When the port mode is GPON, replace the optical module with a GPON&XGS-PON combo optical module. The GPON ONU can work normally.
- When the port mode is GPON&XGS-PON combo mode and the optical module is GPON&XGS-PON combo optical module, the port can work normally only when there are XG(S) ports RTU.

4. Add the XG(S)-PON ONU or replace the GPON ONU with the XG(S)-PON ONU.

NOTE

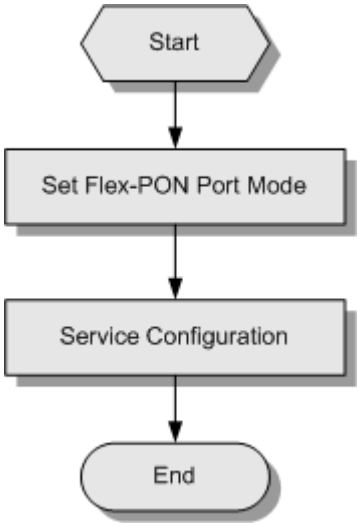
The XGS-PON Combo optical module can be connected to GPON ONU, XG-PON ONU and XGS-PON ONU.

5. Configure services. For details, see the **Flex-PON Configuration Guide**.
6. The evolution is complete.

3.6.3.8 Flex-PON Configuration Guide

Figure 1 shows the flowchart for configuring the Flex-PON service. After setting the Flex-PON port mode, you can configure services according to the configuration cases in the corresponding scenario.

Figure 3-30 Flowchart for configuring the Flex-PON service



Configuration Procedure

Step 1 Set the port mode of the Flex-PON.

 Copyright © Huawei Technologies Co., Ltd.

On the OLT, run the **port flex-mode** command to set the port mode of the Flex-PON.

NOTE

The configured Flex-PON port mode must be the same as the optical module type. Otherwise, the cannot work properly.

For example, set port 0 of the Flex-PON to work in xgs/g combo mode and set the multiplexing mode to WDM.

```
huawei(config-if-gpon-0/1)#port flex-mode
{ all<K>|portid<U><0,15> }:0
{ combo-pon<K>|gpon<K>|xg-pon<K>|xgs-pon<K> }:combo-pon
{ xg/g<K>|xgs/g<K> }:xgs/g
{ <cr>|auto<K>|wdm<K> }:wdm
```

Step 2 Service Configuration. For details, refer to the <Configuration Guide>.

NOTE

After the GPON is switched to the XG(S)-PON, pay attention to the following points:

- For the evolution from GPON to XG-PON Combo or XGS-PON Combo, if the GPON ONU is not replaced after the evolution, the GPON ONU configuration remains unchanged.
- For the evolution from GPON to XG-PON and XGS-PON, the XG-PON ONU or XGS-PON ONU needs to be configured because the GPON ONU is replaced with the XG-PON ONU or XGS-PON ONU after the evolution.

- For an XG(S)-PON FTTH network, configure services according to the XG(S)-PON FTTH Configuration Guide.
  - For carrier customers, visit: **XG(S)-PON FTTH Configuration Guide**
  - For enterprise customers, visit: **XG(S)-PON FTTH Configuration Guide**
- For the XG(S)-PON FTTB/C networking, configure services according to the XG(S)-PON FTTB/C Configuration Guide.
  - For carrier customers, visit: **XG(S)-PON FTTB/C Configuration Guide**
  - For enterprise customers, visit: **XG(S)-PON FTTB/C Configuration Guide**

----End

3.7 XG(S)-PON Reference Standards and Protocols

The following lists standards and protocols:

- ITU-T G.987: ITU-T Recommendation G.987 (2010), G.987 definitions, abbreviations and acronyms
- ITU-T G.987.1: ITU-T Recommendation G.987.1 (2009), 10 Gigabit-capable Passive Optical Network (XG-PON): General Requirements
- ITU-T G.987.2: ITU-T Recommendation G.987.2 (2009), 10 Gigabit-capable Passive Optical Network (XG-PON): Physical Media Dependent (PMD) layer specification
- ITU-T G.987.3: XG-PON TC Layer Specification
- ITU-T G.987.4: XG-PON Reach Extension Specification
- ITU-T G.988: ITU-T Recommendation G.988 (2010), ONU management and control interface specification (OMCI)

 Copyright © Huawei Technologies Co., Ltd.

- ITU-T G.9807.1: ITU-T Recommendation G.9807.1 (2016), 10-Gigabit-capable symmetric passive optical network (XGS-PON)

4 50G-PON

50G-PON is a next-generation high-speed passive optical network (PON) technology evolved from 10G PON. It is a key technology for implementing the 10G optical access network base, ensuring sustainable evolution of ultra-large 10G PON networks, and providing users with higher-speed broadband access capabilities.

4.1 50G-PON Feature Updates

4.2 PON Technology Evolution

4.3 Introduction to 50G-PON

4.4 50G-PON System Architecture

4.5 50G-PON Working Principles

4.6 Key 50G-PON Technologies

4.7 50G-PON Security Policies

4.8 50G-PON Evolution Solution

4.9 50G-PON Service Configuration

4.10 50G-PON Standards and Protocols

4.1 50G-PON Feature Updates

Product Version	Change Description
V100R023C10	First version to support this feature.

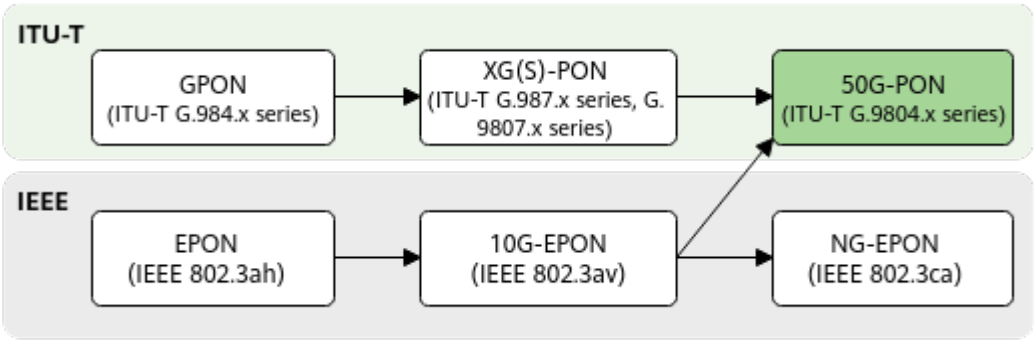
4.2 PON Technology Evolution

PON Technology Evolution Roadmap

Figure 4-1 shows the PON technology evolution roadmap.

 Copyright © Huawei Technologies Co., Ltd.

Figure 4-1 PON technology evolution roadmap



NOTE

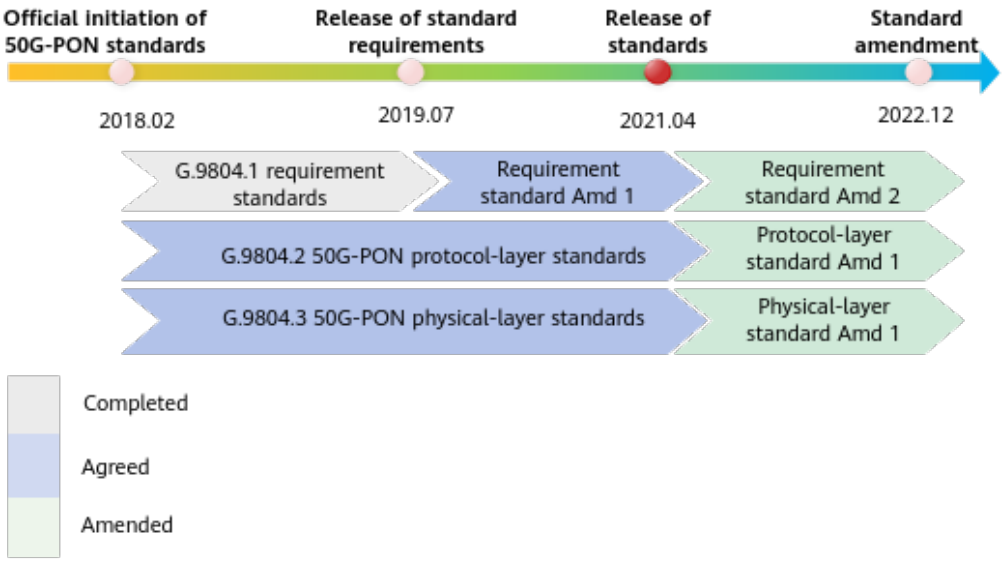
XG(S)-PON includes two technologies: XG-PON and XGS-PON.

Progress of 50G-PON Standards

After formulating the XG-PON standard for 10G PON, standard organizations have started the research on the next-generation PON technology. Based on 10 Gbit/s wavelength superposition, standard organizations started NG-PON2 standard research in 2011 and completed standard formulation in 2015. However, because tunable optical components are costly and relevant systems are less mature, NG-PON2 was rarely used in commercial deployment and may be skipped in the future. At the same time, ITU-T started research on subsequent evolution technologies, released the next-generation high-speed PON technology white paper, and studied the technical possibilities of next-generation high-speed PON access. Compared with the multi-wavelength multiplexing solution, single-wavelength 50G-PON has more potential to become the mainstream industry standard for next-generation optical access networks after 10G PON.

Figure 4-2 shows the progress of 50G-PON standards.

Figure 4-2 Progress of 50G-PON standards



 Copyright © Huawei Technologies Co., Ltd.

- In 2018, ITU/FSAN started the formulation of the single-wavelength 50G-PON standard named G.Higher Speed PON (G.HSP).
- In 2019, the general requirement of 50G-PON G.9804.1 was released, which specifies the single-wavelength time division multiplexing (TDM) PON architecture and upstream and downstream rate combinations. In addition, 50G-PON must support coexistence with and smooth evolution from existing 10G PON and optical distribution networks (ODNs).
- In 2021, ITU-T officially released the 50G-PON series standards, including general requirement amendment G.9804.1 AMD1, general protocol-layer standard G.9804.2, and physical-layer standard G.9804.3. This marks the standardization of basic 50G-PON functions, laying a foundation for the next-step product and solution R&D.
- In 2022, ITU-T amended the 50G-PON series standards, and released the general requirement standard G.9804.1 AMD2, general protocol-layer standard G.9804.2 AMD1, and physical-layer standard G.9804.3 AMD1.

4.3 Introduction to 50G-PON

Why Is 50G-PON Required?

As video becomes a popular broadband service and PON coverage expands from home broadband to industries such as telemedicine, smart manufacturing, and factory and mine communication, access networks need to provide a higher bandwidth, lower delay, with less packet loss and jitter to ensure service quality and user experience. For example, to achieve premium user experience, the VR service requires a bandwidth higher than 1 Gbit/s and a round-trip time (RTT) delay no more than 5 ms. For another example, telemedicine requires an end-to-end delay lower than 50 ms and jitter less than 200 μs.

Compared with 10G PON, 50G-PON increases the bandwidth fivefold and features a low delay, low jitter, and high reliability. Based on the high bandwidth, low delay, and high computing power, 50G-PON can provide premium service experience and meet new application requirements in fields such as digital homes, industry digital transformation, and city digital infrastructure in the 10G era.

Differences in PON Technical Specifications

The following table lists the differences between PON technical specifications.

NOTE

This document describes only the PON technology specifications and does not represent the capabilities of a product. The specific capabilities of a product depend on the actual feature specifications.

 Copyright © Huawei Technologies Co., Ltd.

Item	Wavelength Range	Center Wavelength	Maximum Line Rate	Frame Structure
GPON	• Downstream: 1480–1500 nm • Upstream: 1290–1330 nm	• Downstream: 1490 nm • Upstream: 1310 nm	• Downstream: 2.488 Gbit/s • Upstream: 1.244 Gbit/s	GEM
XG-PON	• Downstream: 1575–1580 nm • Upstream 1260–1280 nm	• Downstream: 1577 nm • Upstream: 1270 nm	• Downstream: 9.953 Gbit/s • Upstream: 2.488 Gbit/s	XGEM
XGS-PON	• Downstream: 1575–1580 nm • Upstream: 1260–1280 nm	• Downstream: 1577 nm • Upstream: 1270 nm	• Downstream: 9.953 Gbit/s • Upstream: 9.953 Gbit/s	XGEM
Asymmetric 50G-PON	• Downstream: 1340–1344 nm • Upstream: 1284–1288 nm	• Downstream: 1342 nm • Upstream: 1286 nm	• Downstream: 49.7664 Gbit/s • Upstream: 24.8832 Gbit/s	XGEM
Symmetric 50G-PON	• Downstream: 1340–1344 nm • Upstream: 1284–1288 nm	• Downstream: 1342 nm • Upstream: 1286 nm	• Downstream: 49.7664 Gbit/s • Upstream: 49.7664 Gbit/s	XGEM

4.4 50G-PON System Architecture

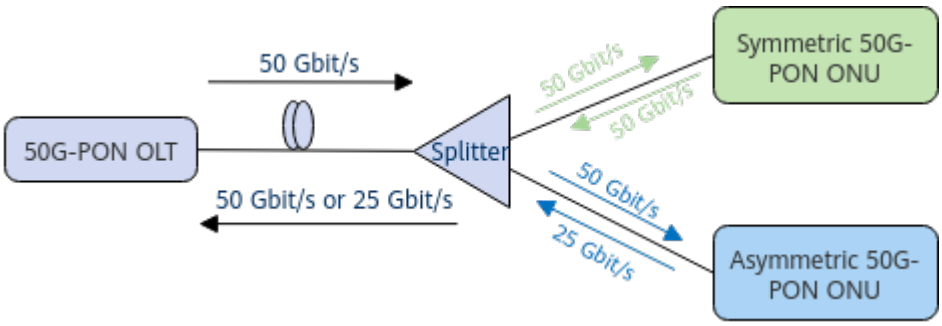
The 50G-PON system architecture complies with the point-to-multipoint (P2MP) architecture of a standard PON network. The physical layer supports the following combinations of upstream and downstream rates:

- Symmetric 50G-PON: 50 Gbit/s downstream and 50 Gbit/s upstream
- Asymmetric 50G-PON: 50 Gbit/s downstream and 25 Gbit/s upstream

To protect customers' investment, 50G-PON devices can share the same ODN with existing PON devices, and allow installed-base PON users to be upgraded smoothly and seamlessly.

 Copyright © Huawei Technologies Co., Ltd.

Figure 4-3 50G-PON system architecture



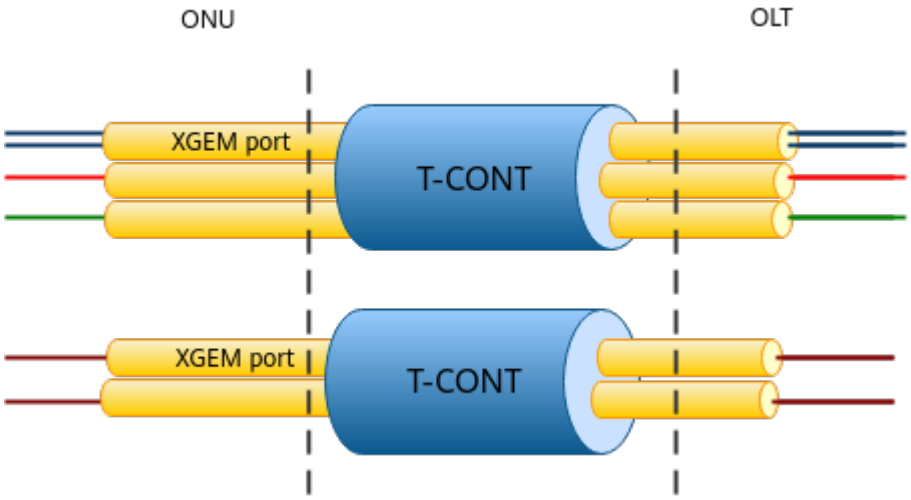
4.5 50G-PON Working Principles

4.5.1 Basic Concepts

Service Multiplexing

XGEM ports and T-CONTs divide a PON network into virtual connections to implement service multiplexing. **Figure 4-4** shows the service multiplexing principles of a 50G-PON system.

Figure 4-4 Service multiplexing principles of a 50G-PON system



XGEM Port

An XGEM port identifies a virtual service channel between an OLT and ONU in a 50G-PON system, that is, a channel that carries service flows. An XGEM port is similar to a virtual path identifier (VPI) or a virtual channel identifier (VCI) in a virtual connection in asynchronous transfer mode (ATM).

- Each XGEM port is identified by a unique XGEM port ID.
- XGEM port IDs are allocated by an OLT based on 50G-PON ports.
- One XGEM port can carry one or more types of services.

 Copyright © Huawei Technologies Co., Ltd.

T-CONT

A T-CONT is the basic control unit of the upstream service flows in a 50G-PON system. It is the carrier for carrying services in the 50G-PON upstream direction. All XGEM ports must be mapped to T-CONTs to transmit data upstream based on dynamic bandwidth assignment (DBA) by the OLT.

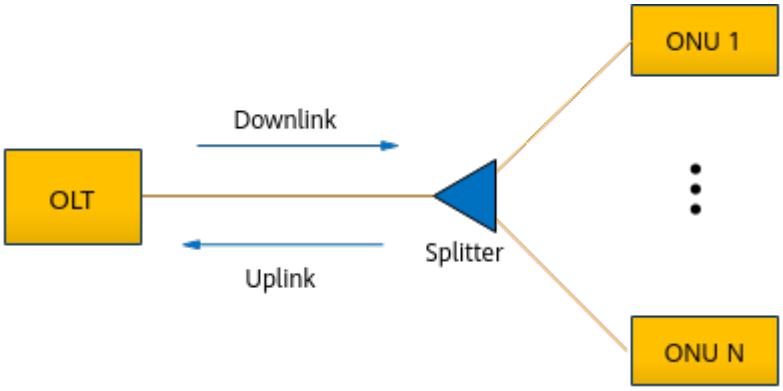
- A T-CONT can carry one or more XGEM ports based on user configurations.
- A T-CONT is identified uniquely by Alloc-ID.
- Alloc-IDs are allocated by an OLT based on 50G-PON ports.
- One ONU supports multiple T-CONTs configured for various service types.

4.5.2 System Transmission Principles

System Transmission Principles

50G-PON uses wavelength division multiplexing (WDM) to transmit bidirectional data over the same fiber but different wavelengths in the upstream and downstream directions on the same ODN, as shown in **Figure 4-5**.

Figure 4-5 System transmission principles



- Data is broadcast in the downstream direction.
- Data is transmitted in time division multiple access (TDMA) mode by ONUs in the upstream direction in different timeslots.

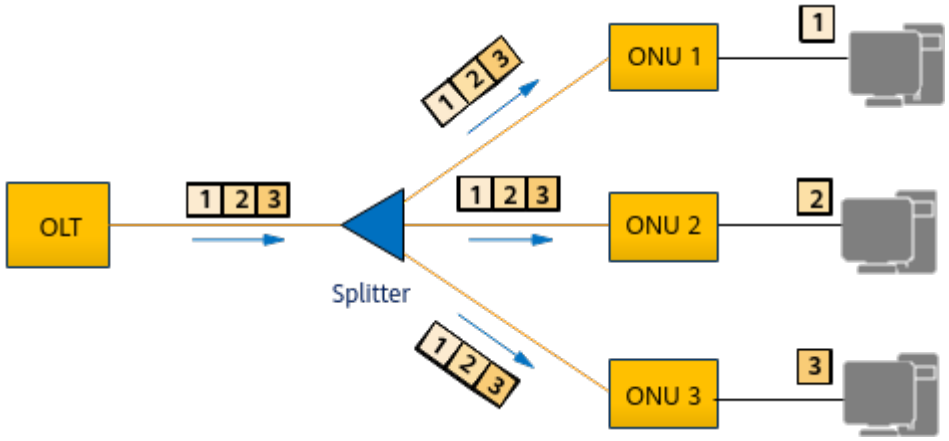
4.5.3 Downstream Working Principles

Downstream Transmission Principles

Figure 4-6 shows the downstream transmission principles.

Copyright © Huawei Technologies Co., Ltd.

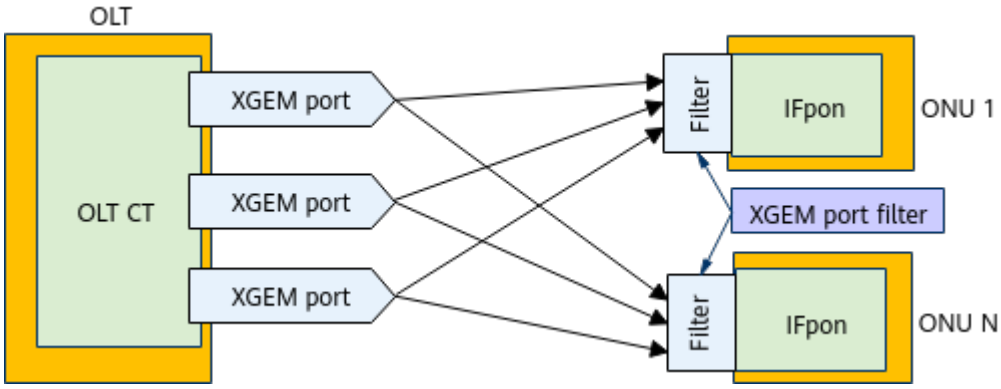
Figure 4-6 Downstream transmission principles



In the downstream direction, data is broadcast by an OLT to all ONUs. Each ONU accepts only data intended for it and discards other data.

Service Flow Forwarding Process

Figure 4-7 Service flow forwarding process in the downstream direction



- An OLT channel termination (CT) multiplexes XGEM frames on a transmission medium by using XGEM port IDs as keys, so as to identify XGEM frames belonging to different downstream logical connections.
- Each ONU filters the downstream XGEM frames by its XGEM port ID and processes only the XGEM frames intended for it.
- One XGEM port can be used to carry XGEM frames to more than one ONU.

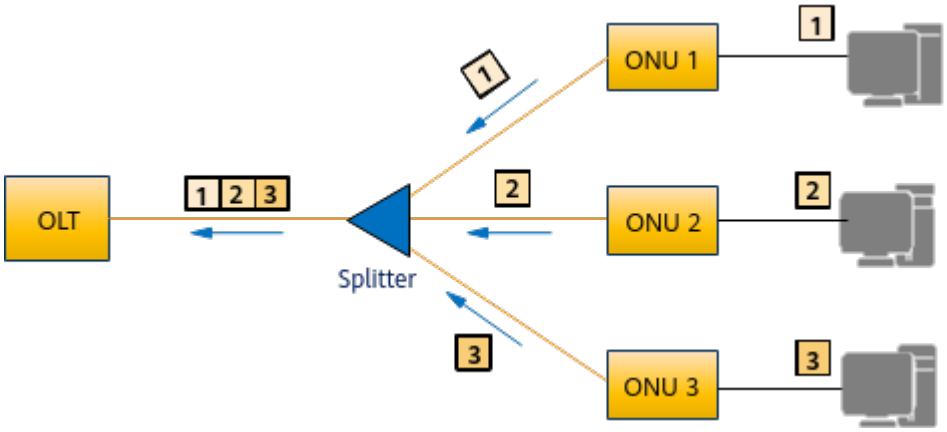
4.5.4 Upstream Working Principles

Upstream Transmission Principles

Figure 4-8 shows the upstream transmission principles.

 Copyright © Huawei Technologies Co., Ltd.

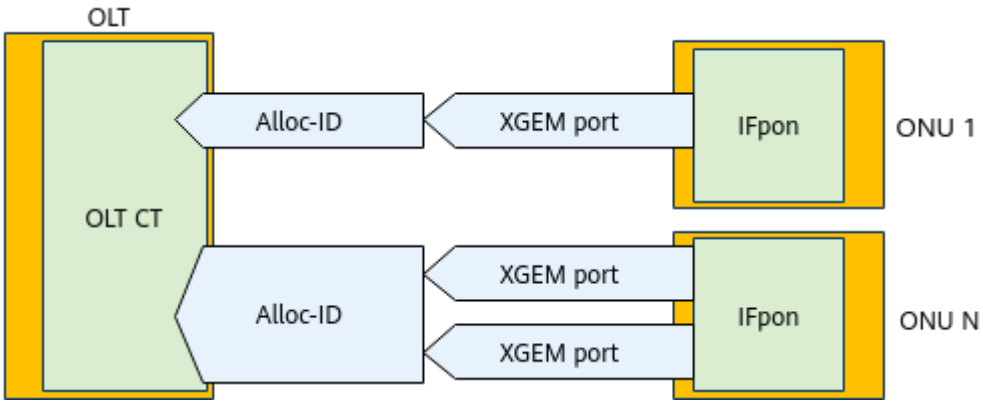
Figure 4-8 Upstream transmission principles



In the upstream direction, an ONU sends data to an OLT in an allocated timeslot in TDMA mode. Such transmission ensures that all ONUs send data in a permitted sequence, and prevents upstream data conflict.

Service Flow Forwarding Process

Figure 4-9 Service flow forwarding process in the upstream direction



- The OLT CT grants upstream transmission opportunities or upstream bandwidth allocations to the traffic-bearing entities within subtending ONUs.
- Within an ONU, the traffic-bearing entity that receives the upstream bandwidth allocations is identified by its allocation ID (Alloc-ID).
- The bandwidth assigned to different Alloc-IDs is multiplexed in the time specified by the OLT CT in the bandwidth mapping for downstream transmission.
- In each bandwidth assignment, an ONU uses the XGEM port ID as the multiplexing key to identify XGEM frames belonging to different upstream logical connections.

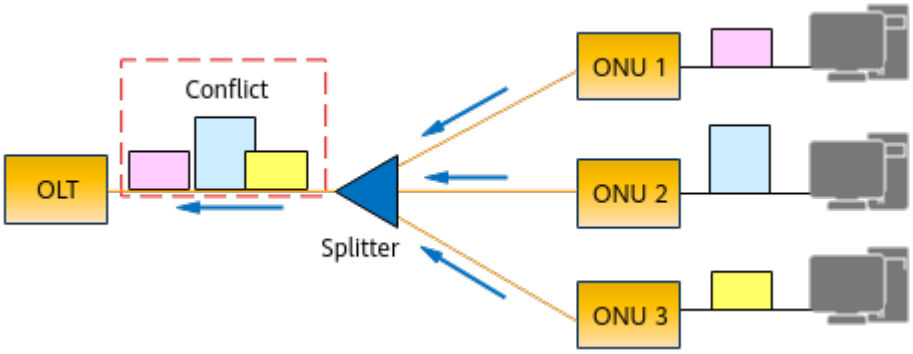
4.6 Key 50G-PON Technologies

 Copyright © Huawei Technologies Co., Ltd.

4.6.1 Ranging

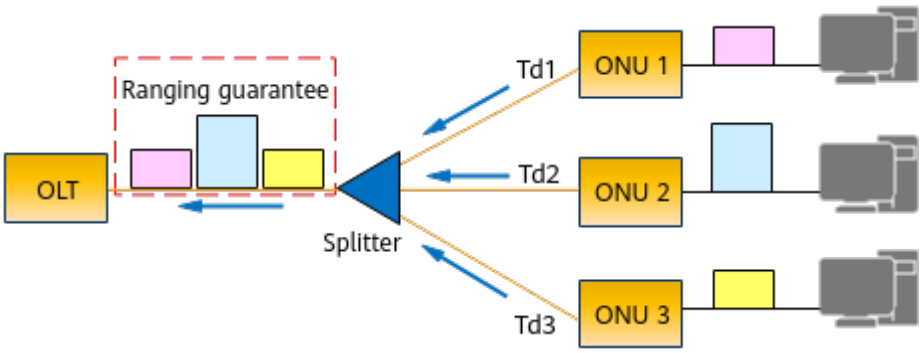
The physical distances from optical network units (ONUs) to an optical line terminal (OLT) differ from case to case. The round trip delays (RTDs) between an OLT and ONUs also vary as time goes by and the environment changes. Therefore, collisions may occur when ONUs send data in TDMA mode (in this mode, only one of the ONUs connected to an OLT PON port sends data at a moment), as shown in **Figure 4-10**.

Figure 4-10 Cell transmission without ranging



To avoid collisions, ranging is triggered when an ONU is registered for the first time. The OLT measures the RTD of each ONU in the ranging process and adds an equalization delay (EqD, represented by the parameter Td) for each ONU to ensure that the logical transmission distances of all ONUs connected to the same OLT PON port are the same. This prevents collisions during upstream transmission.

Figure 4-11 Cell transmission with ranging



NOTE

In the ranging process, the OLT must open a window and pause upstream transmission channels of other ONUs.

4.6.2 Burst Optical/Electrical Technology

Working Principles of Burst Optical/Electrical Technology

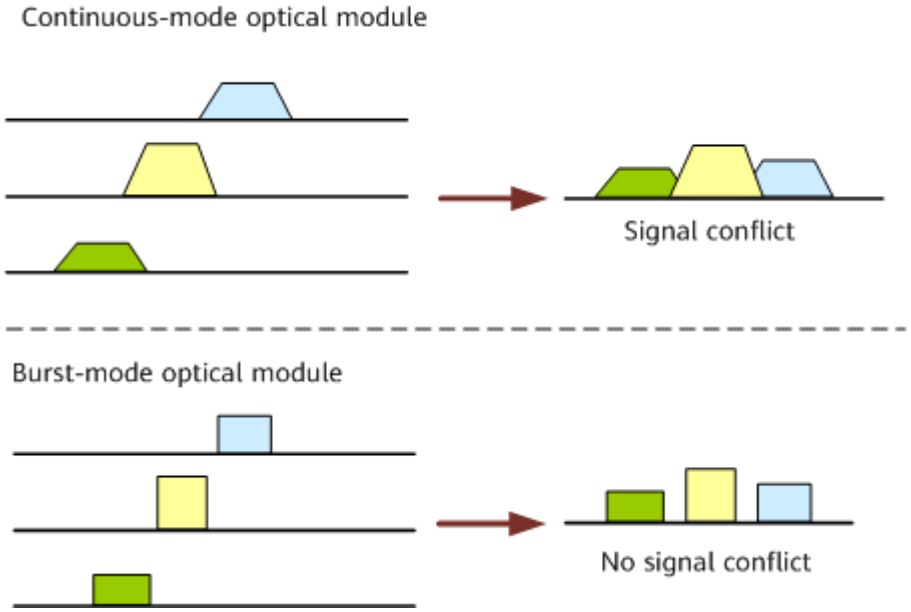
In the upstream direction, 50G-PON works in TDM mode. Each ONU can send data only in the permitted timeslot. In the timeslot that does not belong to an

Copyright © Huawei Technologies Co., Ltd.

ONU, the optical module of the ONU must be disabled from sending signals. In this way, other ONUs can work properly. An OLT receives signals sent upstream from each ONU in burst mode by timeslot. To ensure the normal operation of the system:

- The optical module of an ONU must support burst transmission (as shown in **Figure 4-12**).
- The optical module of an OLT must support burst receiving (as shown in **Figure 4-13**).

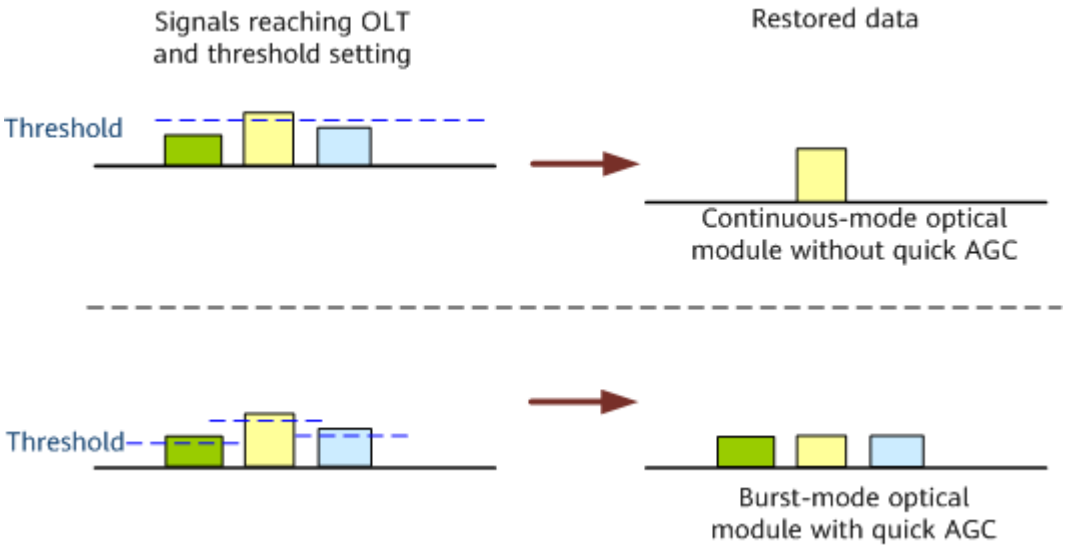
Figure 4-12 Burst transmission by an ONU



Ranging can be implemented to prevent cells transmitted by different ONUs from conflicting with each other on the OLT. However, the ranging accuracy is ± 1 bit and the cells transmitted by different ONUs have a protection time of several bits (not a multiple of 1 bit). If the ONU optical modules do not support the burst receive and transmit function, the transmitted signals overlap and distortion occurs.

 Copyright © Huawei Technologies Co., Ltd.

Figure 4-13 Burst receiving by an OLT



- The distance from each ONU to an OLT varies and therefore the optical signal attenuation varies for each ONU. As a result, the power and level of packets received by an OLT vary with time slots.
- If the optical module of an OLT does not support quick processing of burst optical power changes, signals may be incorrectly restored. (The signals are considered valid only when the level is higher than the threshold. If the level is lower than the threshold, the signals cannot be restored correctly.) The automatic gain control (AGC) function enables an OLT to dynamically adjust the threshold of the receive optical power according to the strength of the received optical signals to ensure that the signals of all ONUs can be correctly restored.

NOTE

In the 50G-PON downstream direction, an OLT broadcasts data to all ONUs. Therefore, the optical module of an OLT must support continuous transmission, and that of an ONU must support continuous receiving. The burst transmission and receiving functions are not required for the optical modules in the downstream direction.

Single-Frame Multi-Burst Technology

Because a PON system uses P2MP physical connections, the OLT needs to properly allocate receiving timeslots to all ONUs. For a specific ONU, the upstream timeslots are limited. The period after the end of an upstream timeslot to the start of the next upstream timeslot belongs to other ONUs. In this period, the ONU cannot send any data upstream, and caches data locally to transmit data at the next timeslot. Therefore, the upstream timeslot interval has a great impact on the upstream delay of ONUs.

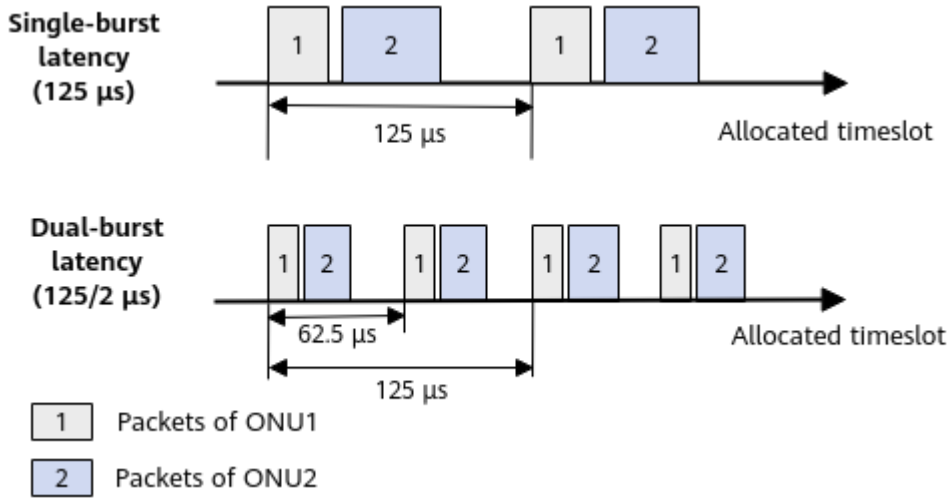
In an ITU TDM PON system, the upstream timeslot allocation packet of an ONU is located in the bandwidth (BW) map with a fixed overhead in the frame header of each downstream frame, and each BW map corresponds to a 125 μs upstream timeslot allocation for a downstream frame.

Figure 4-14 shows the mapping between the number of burst timeslots obtained by an ONU in each frame and the maximum delay of the upstream data. If ONU1

 Copyright © Huawei Technologies Co., Ltd.

has only one upstream transmission timeslot in every 125 μs, the transmission interval can reach 125 μs, and the maximum upstream delay is 125 μs. If ONU1 has two upstream transmission timeslots in every 125 μs, the interval between the timeslots is 62.5 μs (125/2), and the upstream delay is reduced by 1/2. By analogy, if ONU1 has N upstream transmission timeslots in every 125 μs, the interval between the timeslots is 125/N μs, and the upstream delay is reduced to 1/N.

Figure 4-14 Principle of the single-frame multi-burst technology (using the single-frame dual-burst technology as an example)



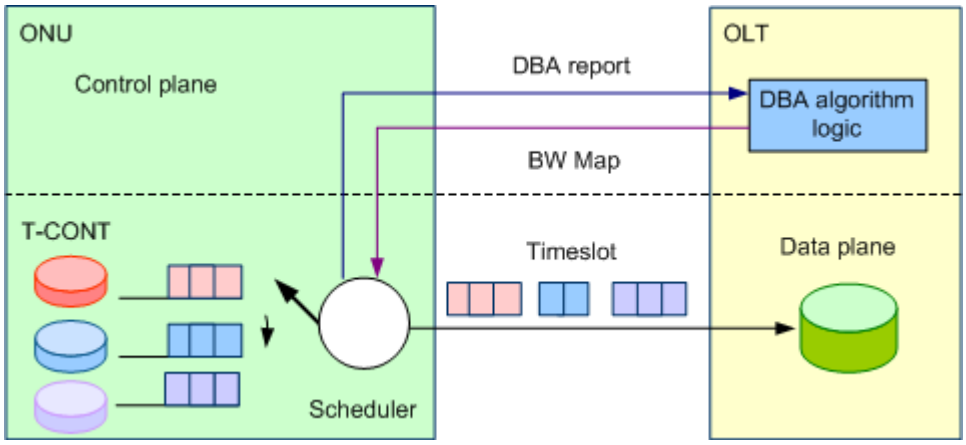
4.6.3 DBA

DBA Principles

DBA is short for dynamic bandwidth assignment. An OLT uses DBA to dynamically adjust the upstream bandwidth assigned to different ONUs to process the burst traffic from ONUs, meeting the ONU upstream bandwidth requirements and improving the utilization of the PON upstream bandwidth.

Figure 4-15 shows the DBA principles.

Figure 4-15 DBA principles



Copyright © Huawei Technologies Co., Ltd.

DBA implementation process:

- The DBA module in an OLT consistently collects DBA reports and uses a DBA algorithm to calculate the upstream bandwidth assigned to each ONU.
- The OLT sends the calculated result to each ONU using a bandwidth (BW) map.
- Each ONU transmits upstream data in burst mode in permitted timeslots defined in the BW map.

Highlights and Applications

DBA offers the following features:

- Based on ONUs' burst upstream service traffic, the OLT dynamically allocates an upstream bandwidth to each ONU in real time, improving upstream bandwidth utilization on PON ports.
- More users are supported on a PON port.
- Higher service bandwidths with burst requirements are supported than those before DBA is applied.

4.6.4 FEC

50G-PON increases the line rate, but the receiver sensitivity decreases, and transceiver performance needs to be improved to reuse massive existing ODNs. To lower the specification requirements for high-speed optical components, 50G-PON uses the low density parity check (LDPC) (17280, 14592) encoding and decoding solution to perform forward error correction (FEC), with an encoding efficiency about 84.44%.

Principles

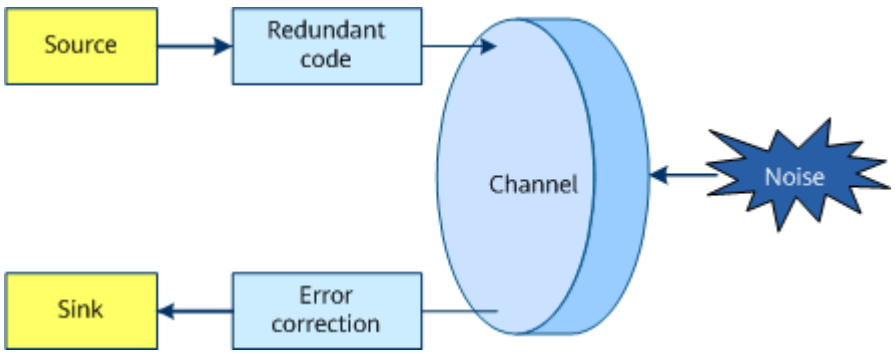
FEC is a data encoding technology that enables a data receiver to check for bit errors during transmission based on the encoding scheme.

Redundant codes are added to signals on the transmitter as error-correction codes (ECCs). Then, the receiver checks the signals for errors based on the ECCs and corrects an error if there is any.

Common FEC technologies include the Hamming code, RS code, LDPC code, convolutional code, and the like.

Figure 4-16 shows FEC principles.

Figure 4-16 FEC principles



Copyright © Huawei Technologies Co., Ltd.

The FEC algorithms used by 10G PON and 50G-PON comply with ITU-T standards, as shown in **Table 4-1**.

Table 4-1 FEC algorithms used by 10G PON and 50G-PON

PON Technology	Upstream	Downstream
XG-PON	RS (248, 232)	RS (248, 216)
XGS-PON	RS (248, 216)	RS (248, 216)
50G-PON	LDPC (17280, 14592)	LDPC (17280, 14592)

NOTE

The LDPC algorithm is classified into hard value input and soft value input based on whether the output data is decided. The hard value data is a 0 and 1 bit sequence of the signals after equalization decision and is input to the LDPC decoder for decoding. The soft value data is the original data that is not decided, and the log-likelihood ratio (LLR) of the input signals can be obtained, thereby improving the error correction performance.

Features and Applications

- Does not require retransmission and provides a high real-time performance.
- Requires an additional bandwidth. Users must balance the transmission quality and bandwidth.
- Checks and corrects errors after data is transmitted to the peer end, and does not apply to services for which retransmission is enabled.
- Applies to data transmission when the network quality is poor.
- Applies to services that have a low requirement on delay (The delay is large if retransmission is configured for services.).

4.7 50G-PON Security Policies

4.7.1 Line Encryption

Context

Line encryption technologies are required to prevent the data theft risk.

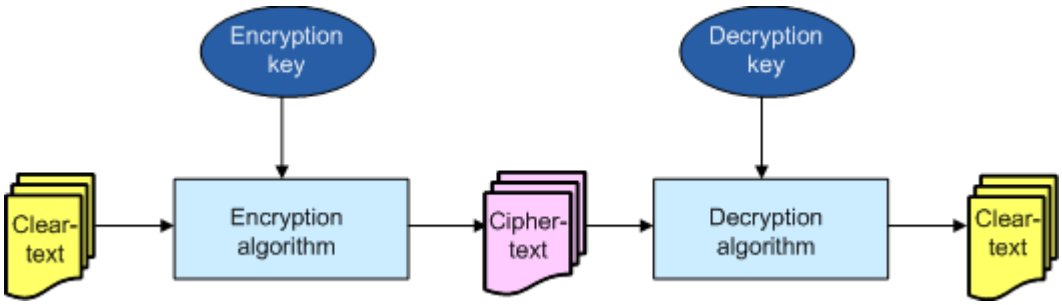
50G-PON line encryption supports bidirectional encryption in the upstream and downstream directions to ensure secure bidirectional data transmission.

Working Principles

Figure 4-17 shows the working principles of 50G-PON line encryption, in which clear-text packets are encrypted into cipher-text packets using encryption algorithms (AES-128, AES-256) to improve transmission security.

 Copyright © Huawei Technologies Co., Ltd.

Figure 4-17 Line encryption process



To prevent the risks caused by the cracking of the encryption key, the key update technology is used to periodically update the encryption key to improve the security.

Features and Applications

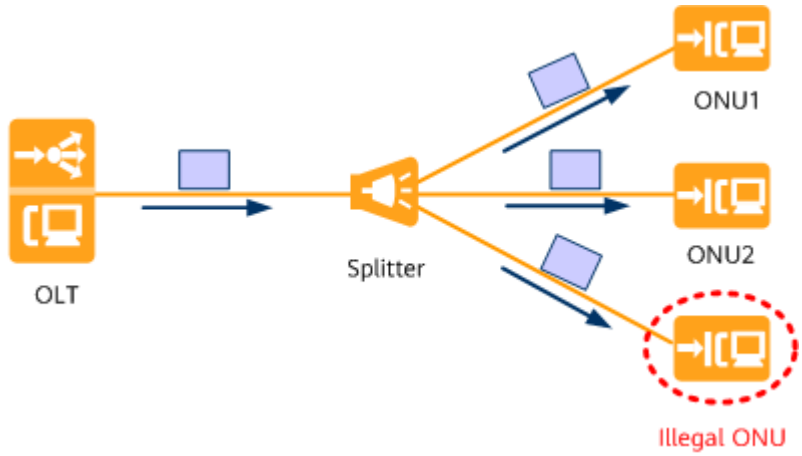
- The line encryption algorithms neither increase the overhead nor require additional bandwidth.
- The line encryption algorithms do not increase the transmission delay.
- It is recommended that you enable the line encryption function in scenarios that have high requirements on line security.

4.7.2 ONU Authentication

Context

In the P2MP PON system architecture, downstream data is broadcast to all ONUs, as shown in **Figure 4-18**. As a result, illegal ONUs may also have the opportunity to receive data.

Figure 4-18 Downstream data transmission process



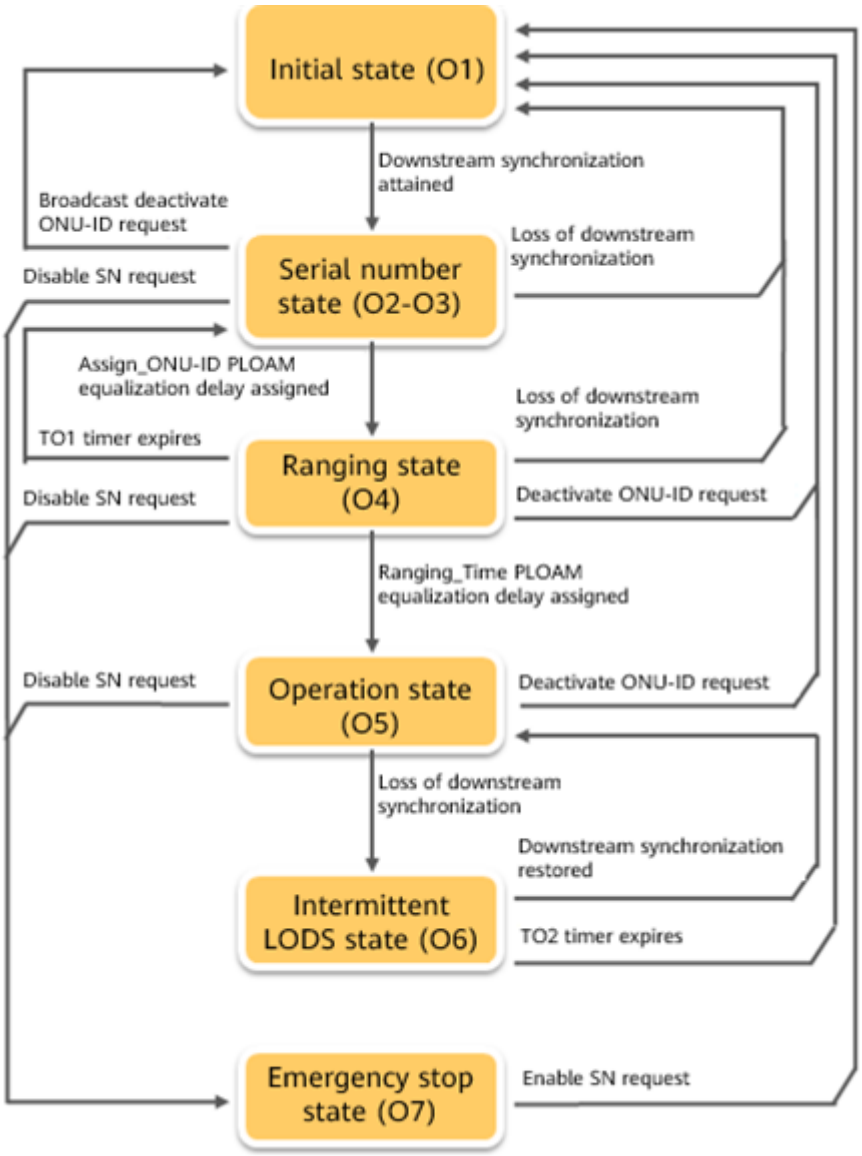
To solve this problem, a PON system uses authentication technology to ensure the validity of connected ONUs.

Copyright © Huawei Technologies Co., Ltd.

Authentication Process

Figure 4-19 shows the 50G-PON authentication process, involving six states and two timers.

Figure 4-19 Authentication process



Six states:

- O1: Initial state
- O2-3: Serial number state
- O4: Ranging state
- O5: Operation state
- O6: Intermittent LODS state
- O7: Emergency stop state

 Copyright © Huawei Technologies Co., Ltd.

Two timers:

- TO1 timer
- TO2 timer

4.7.3 Continuous-Mode ONU Detection

Overview

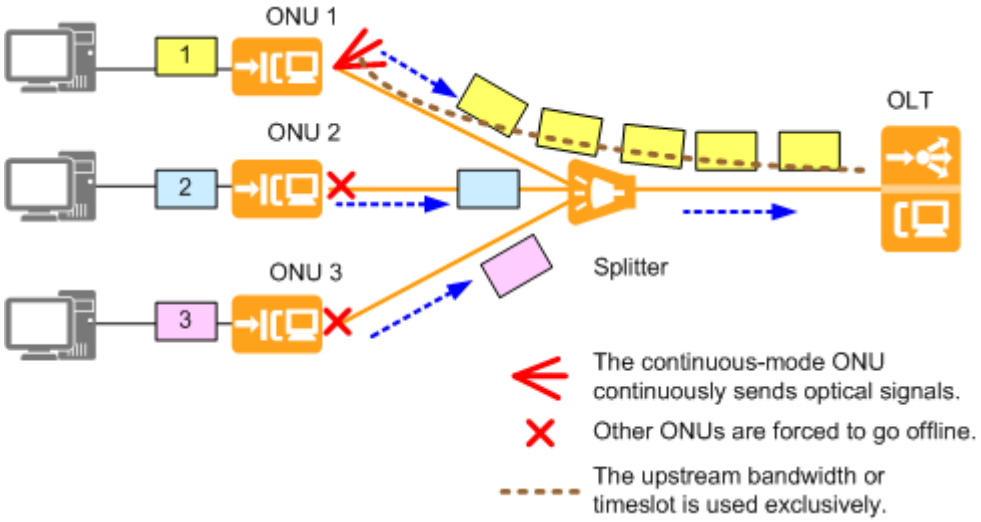
PON networks use the P2MP network architecture. They use TDMA in the upstream direction. ONUs must send optical signals upstream at the timeslots allocated by an OLT to prevent data conflict.

A continuous-mode ONU is also called a continuous-mode rogue ONU. It continuously sends optical signals to the upstream direction.

A continuous-mode ONU adversely affects the system as follows:

- If this ONU is online, some or all ONUs connected to the same PON port may go offline or frequently go offline and online.
- If this ONU has not been configured, other ONUs connected to but not configured on the same PON port will not be discovered by the OLT.

Figure 4-20 Rogue ONU



Prerequisites for Rogue ONU Detection

- By checking and analyzing the optical signals in the upstream direction, an OLT can identify and isolate rogue ONUs from non-malicious users, but cannot identify ONUs that are intentionally damaged or that are incompliant with standard requirements. How to handle these types of ONUs is not within the scope of this document.
- A continuous-mode ONU can be detected only when it can parse downstream PLOAM messages and correctly respond to the messages.

Copyright © Huawei Technologies Co., Ltd.

- An OLT can quickly locate a continuous-mode ONU connected to a PON port only when the ONU supports standard PLOAM messages and the optical module of the ONU is correctly controlled.

Continuous-Mode ONU Detection

Continuous-mode ONU detection, also called rogue ONU detection, is used to detect continuous-mode ONUs and isolate them to ensure proper system running. The continuous-mode ONU detection process involves three stages, checking, detection, and isolation.

- Checking: The OLT periodically checks whether a continuous-mode ONU is connected to a PON port. The checking process determines whether there is a continuous-mode ONU, but cannot locate any continuous-mode ONU.

  The OLT opens an empty window in the upstream direction to detect whether any ONU sends optical signals upstream in this empty window. If the OLT receives any optical signals, it then goes to the detection stage to locate the ONU that sends the optical signals.
- Detection: The OLT locates a continuous-mode ONU.

  The OLT issues commands to ONUs in turn to instruct the optical module of each ONU to send optical signals upstream, and checks whether any ONU sends optical signals at a time not within the authorized timeslot. If an ONU sends optical signals at a time not within the authorized timeslot and causes other online ONUs to go offline, this ONU is a continuous-mode ONU. In the detection process, the OLT checks all ONUs connected to a PON port to detect all continuous-mode ONUs.
- Isolation: The OLT issues a command to power off the optical module of a continuous-mode ONU, preventing this ONU from adversely affecting other ONUs connected to the same PON port.

  Once the optical module of an ONU is disabled from sending optical signals upstream by the OLT, the effect is permanent unless the OLT issues another command to enable the upstream optical signal transmission. That is, the upstream optical signal transmission will not be enabled again even after the ONU is reset or powered off and restarted. This mechanism ensures that a continuous-mode ONU is completely isolated.

NOTE

By default, an OLT only checks for rogue ONUs and does not automatically detect or isolate rogue ONUs.

4.8 50G-PON Evolution Solution

4.8.1 50G-PON Network Evolution Overview

PON networks that have been deployed by operators can smoothly evolve to 50G-PON networks.

Why Smooth Evolution Is Required?

During network evolution, operators attach much importance to reusing live network resources and reducing upgrade and evolution costs. To enable smooth

 Copyright © Huawei Technologies Co., Ltd.

evolution from deployed PON networks to 50G-PON networks and meet networking requirements of different services, deployed PON networks and 50G-PON networks may coexist for a certain period of time. To save the deployment space in equipment rooms, reduce the power consumption of optical access equipment, effectively reuse ODN resources on the live network, and reduce the network construction costs of operators, a proven best practice is to use multi-mode optical transceiver modules in CO equipment. Therefore, during the evolution to 50G-PON, two-step migration can be adopted for PON technology coexistence. First, evolution from PON to 10G PON, and then evolution from 10G PON to 50G-PON. Continuous bandwidth upgrade is achieved while network evolution and costs are considered.

Prerequisites for Smooth Evolution from PON to 50G-PON

50G-PON provides the following features to support smooth evolution from existing PON to 50G-PON:

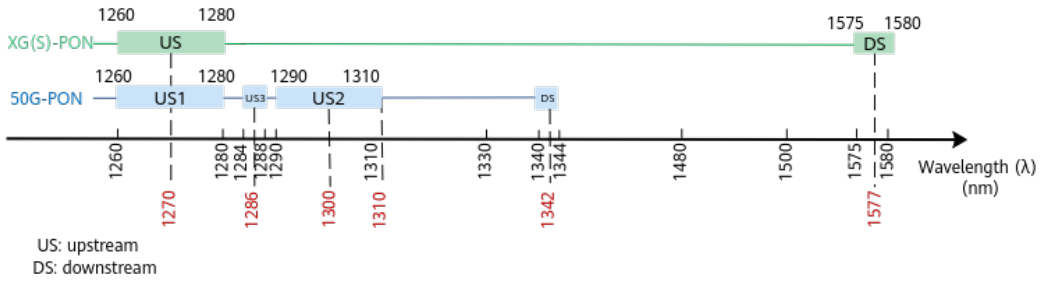
- In the upstream direction, 50G-PON uses the US3 (1284–1288 nm) wavelength option, which does not overlap with the wavelength planning for XG(S)-PON. Therefore, 50G-PON can share an ODN with XG(S)-PON in WDM mode.
- The service provisioning and deployment operations for 50G-PON and existing PON networks are largely the same. Therefore, the network management system and service provisioning system can be reused.

50G-PON Wavelength Option for Coexisting with XG(S)-PON

In the G.9804.1 standard for 50G-PON, a third upstream wavelength frequency band US3 (1284–1288 nm) is added between the existing upstream wavelength frequency bands US1 (1260–1280 nm) and US2 (1290–1310 nm). 50G-PON can use the US3 (1284–1288 nm) wavelength option which does not overlap with the wavelength planning for XG(S)-PON to coexist with XG(S)-PON on the live network in WDM mode.

Figure 4-21 Shows the wavelength distribution of XG(S)-PON and 50G-PON.

Figure 4-21 Wavelength distribution of XG(S)-PON and 50G-PON



4.8.2 Working Principles of 50G-PON Combo PON

The dual-mode Combo PON technology enables PON systems complying with ITU-T and IEEE standards to coexist with and evolve to 50G-PON.

 Copyright © Huawei Technologies Co., Ltd.

4.8.3 Evolution Solution for XGS-PON Networks

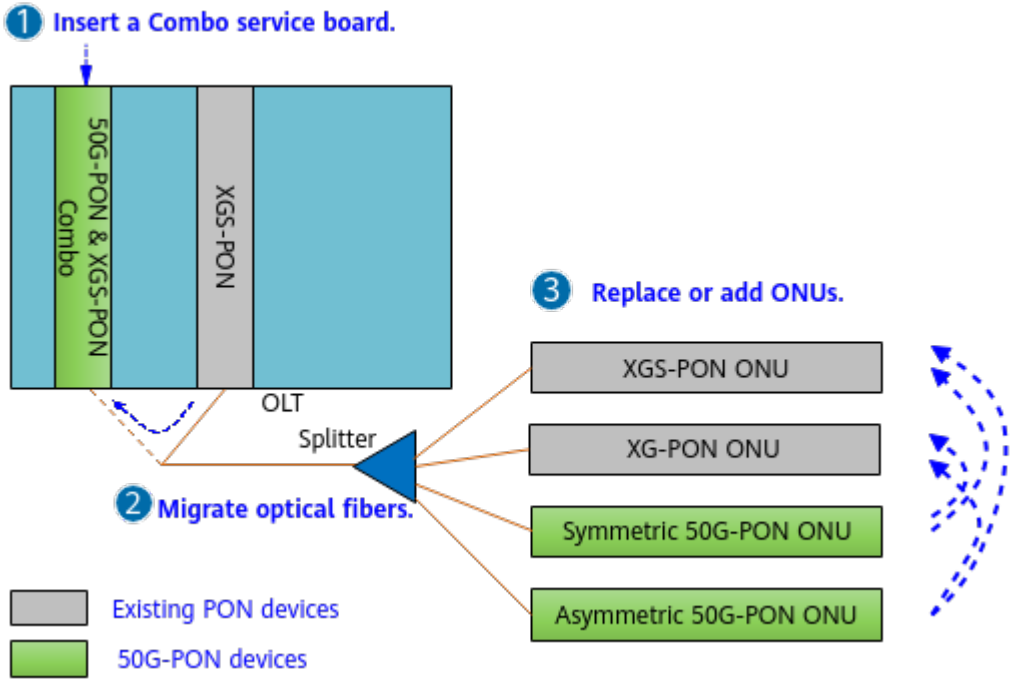
This topic describes how to smoothly upgrade XGS-PON to 50G-PON using the symmetric 50G-PON & XGS-PON dual-mode Combo solution.

50G-PON uses upstream and downstream wavelengths different from those used by XGS-PON. With the WDM technology, 50G-PON and XGS-PON services can be carried over the same optical fiber. Customers only need to replace the boards in OLT service subracks with symmetric 50G-PON & XGS-PON dual-mode Combo boards. In this way, XGS-PON can be smoothly upgraded to symmetric 50G-PON by reusing live ODNs without adding external devices or facilities. This saves equipment room space and construction costs, and improves deployment and O&M efficiency.

Evolution Solution

Add symmetric 50G-PON & XGS-PON dual-mode Combo service boards to implement smooth evolution from XGS-PON to symmetric 50G-PON, as shown in **Figure 4-22**.

Figure 4-22 Evolution from XGS-PON to symmetric 50G-PON



Procedure

1. Add symmetric 50G-PON & XGS-PON dual-mode Combo service boards to an OLT.
2. Migrate the feeder optical fibers from XGS-PON ports to symmetric 50G-PON & XGS-PON dual-mode Combo ports.
3. Add symmetric 50G-PON ONUs or replace existing XG-PON or XGS-PON ONUs on the live network.

Copyright © Huawei Technologies Co., Ltd.

NOTE

- An XGS-PON port can connect XG-PON and XGS-PON ONUs.
- The planned logical distances from XG(S)-PON, symmetric 50G-PON, and asymmetric ONUs to symmetric 50G-PON & XGS-PON dual-mode Combo ports must be the same after ranging and compensation.

4. Configure services. For details, see **4.9 50G-PON Service Configuration**.

4.9 50G-PON Service Configuration

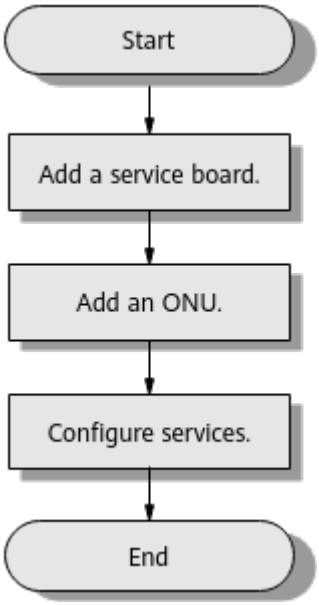
Context

Figure 4-23 shows the process of configuring 50G-PON services. After the 50G-PON Combo mode is configured, you can configure the services according to the configuration cases for the corresponding scenarios.

NOTE

This document uses the 50G-PON Combo service board as an example.

Figure 4-23 50G-PON service configuration



Procedure

Step 1 Add a service board.

In global config mode, run the **board add** *frameid/slotid board-type* command to add a service board.

You can add a service board to an OLT in either of the following ways:

- Manually insert the service board. This method is applicable to the scenario where a service board already exists. You can directly insert the service board into a subrack.

 Copyright © Huawei Technologies Co., Ltd.

- Add a service board offline. This method is applicable to the scenario where a service board is not inserted into a subrack and services need to be pre-configured for the board.

  NOTE

  - After a service board is added successfully, it is in the Failed state. You can still configure and query the data of the service board.
  - After you manually insert a service board into the corresponding slot, the service board status is switched to **Normal** and the configuration data takes effect. The type of the manually inserted service board must be the same as that of the service board added offline; otherwise, the service board cannot run normally.

Step 2 Add an ONU.

You can add an ONU to an OLT in either of the following ways:

- Add an ONU offline. In this scenario, before an ONU is installed, run the **ont add** command on the OLT to manually add the ONU and configure it. After the ONU goes online, the OLT authenticates the ONU and issues configurations to it.
- Add an ONU online. In this scenario, after an ONU is installed, run the **port ont-auto-find** command to enable the ONU auto-find function. The ONU auto-find function is used to periodically check whether a new ONU goes online. After an online ONU is found, the ONU is added and configured.

  NOTE

  - By default, the ONU auto-find function is disabled for an OLT PON port.
  - You can run the **display ont autofind** command to query the ONUs that have been automatically found by a port.

Step 3 Configure the services according to the corresponding configuration guide.

- For a 50G-PON FTTH network, configure services according to the FTTH configuration guide.
  - For carrier customers, visit: **FTTH Configuration Guide**
  - For enterprise customers, visit: **FTTH Configuration Guide**
- For a 50G-PON FTTB/FTTC network, configure services according to the FTTB/FTTC configuration guide.
  - For carrier customers, visit: **FTTB and FTTC Configuration Guide**
  - For enterprise customers, visit: **FTTB and FTTC Configuration Guide**

----End

4.10 50G-PON Standards and Protocols

50G-PON standards are as follows:

- ITU-T G.9804.1 (11/2019): Higher speed passive optical networks – Requirements
- ITU-T G.9804.2 (09/2021): Higher speed passive optical networks – Common transmission convergence layer specification
- ITU-T G.9804.3 (09/2021): 50-Gigabit-capable passive optical networks (50G-PON): Physical media dependent (PMD) layer specification

 Copyright © Huawei Technologies Co., Ltd.

- ITU-T G.9805 (02/2022): Coexistence of passive optical network systems
- ITU-T G.988 (11/2022): ONU management and control interface (OMCI) specification

 Copyright © Huawei Technologies Co., Ltd.

5 Ethernet Networking Applications

5.1 P2P Optical Access
Point-to-point (P2P) Ethernet optical access refers to the P2P FTTX access provided by the P2P Ethernet optical access board and the ONT, which meets the requirements for the application of the next generation access device under the integration of video, voice, and data services.

5.2 NE Cascading
NE cascading refers to a networking mode in which the MA5800 series are directly connected to each other through the ethernet port on the board. NE cascading saves the upstream optical fiber resources of the access node.

5.3 Ethernet OAM
Ethernet Operation, Administration, and Maintenance (OAM) is used to operate, administrate, and maintain Ethernet networks. It covers three features: connectivity fault management (CFM), Ethernet in the first mile (EFM), and performance monitoring (PM).

5.1 P2P Optical Access

Point-to-point (P2P) Ethernet optical access refers to the P2P FTTX access provided by the P2P Ethernet optical access board and the ONT, which meets the requirements for the application of the next generation access device under the integration of video, voice, and data services.

5.1.1 What Is P2P Optical Access

Definition

Point-to-point (P2P) Ethernet optical access is a mode in which P2P Ethernet optical access boards provide GE, 10GE and FE ports and coordinate with downstream devices to implement various optical access solutions for users. The solutions include FTTC/FTTB, FTTH, FTTO, FTTM.

Purpose

P2P Ethernet optical access is specially applicable to residential communities with optical fibers routed. It provides a more flexible FTTx solution, especially

 Copyright © Huawei Technologies Co., Ltd.

integrated services, including video, voice, and data for users. This feature has the following advantages:

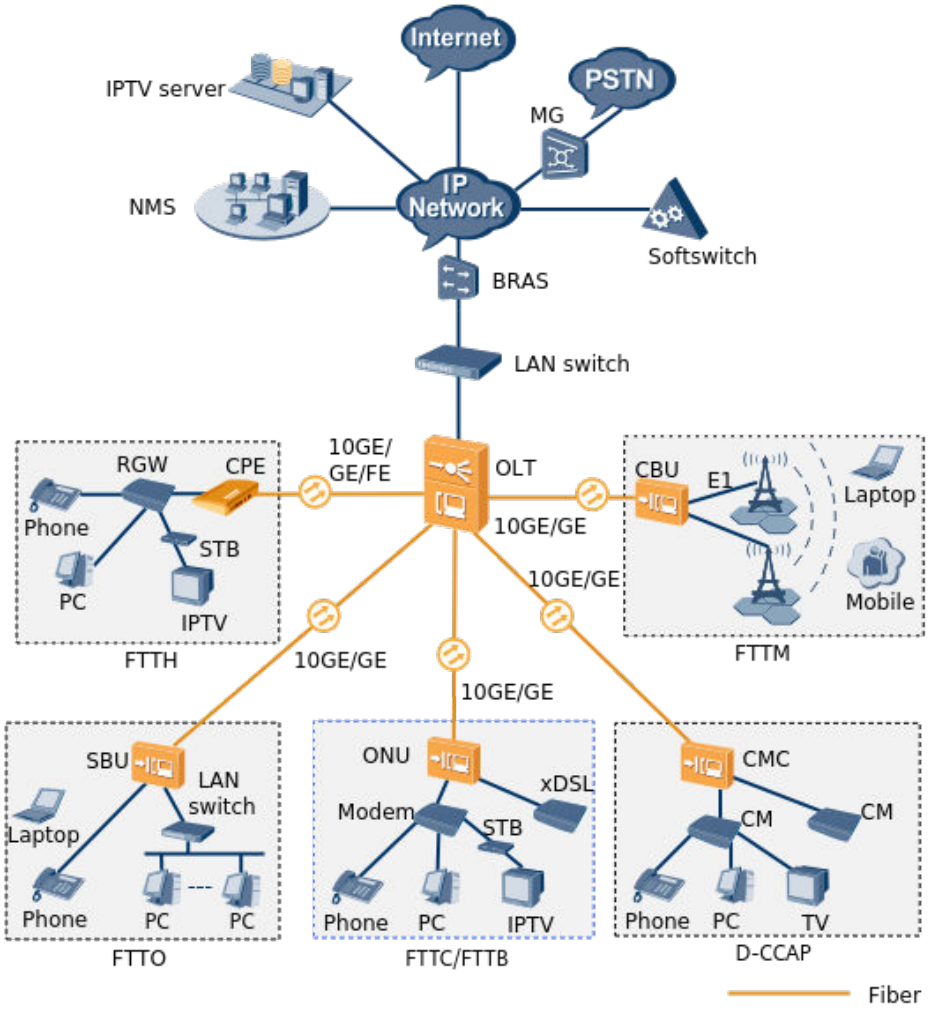
- Higher bandwidth. The FTTH solution implemented through P2P optical access can provide a higher bandwidth for users, meeting the requirements of high-end users.
- Lower costs. P2P Ethernet optical access boards support more cascaded DSLAMs, reducing FTTC/FTTB networking costs.
- Higher reliability. P2P optical access allows higher reliability in the DSLAM cascading scenario through features such as inter-board aggregation, and ring check.
- More flexible scenarios. The P2P Ethernet optical access boards coordinate with a variety of downstream devices (such as the DSLAM, ONT, SBU, and CBU) to implement FTTC/FTTB, FTTH, FTTO, FTTM. An MA5800 configured with the P2P Ethernet optical access board cannot only be directly connected to access terminals but also cascade DSLAMs in order to converge a large number of users.

5.1.2 Networking Application

Figure 5-1 shows the network application in the P2P Ethernet optical access mode.

Copyright © Huawei Technologies Co., Ltd.

Figure 5-1 Network application in the P2P Ethernet optical access mode



To meet the requirements of different scenarios, the OLT works with ONUs of various types to implement network applications in multiple optical access modes, such as FTTC/FTTB, FTTH, FTTO, FTTM.

The FTTx network applications in P2P Ethernet optical access have the following in common: The data, voice, and video signals of terminal users are sent to ONUs, where the signals are converted into Ethernet packets and then transmitted over optical fibers to the OLT through the GE/FE upstream ports of the ONUs. Then, the Ethernet packets are forwarded to the upper-layer IP network through the upstream port of the OLT.

The differences between the networking applications of the P2P Ethernet optical access are as follows:

- FTTH: The OLT is connected to the CPEs at user premises through P2P Ethernet optical access. In this way, gigabit bandwidth is exclusively provided to each household. FTTH is applicable to new apartments or villas in loose distribution. In this scenario, FTTH provides services of higher bandwidth for high-end users.
- FTTB/FTTC: The OLT is connected to DSLAMs in corridors (FTTB) or by the curb (FTTC) through GE P2P Ethernet optical access. The DSLAMs are then

 Copyright © Huawei Technologies Co., Ltd.

connected to user terminals through xDSL. With the aggregation provided by the DSLAMs, one port on the P2P interface board can be connected to a large number of users. FTTB/FTTC is applicable to densely-populated residential communities or office buildings. In this scenario, FTTB/FTTC provides services of certain bandwidth for common users.

- FTTO: The OLT is connected to enterprise SBUs through GE P2P Ethernet optical access. The SBUs are connected to user terminals through FE, POTS, or Wi-Fi. QinQ VLAN encapsulation is implemented on the SBUs and the OLT. In this way, transparent and secure data channels can be set up between the enterprise private networks located at different places, and thus the service data and BPDUs between the enterprise private networks can be transparently transmitted over the public network. FTTO is applicable to enterprise networks. In this scenario, FTTO implements TDM PBX, IP PBX, and private line service in the enterprise intranets.
- FTTM: The OLT is connected to CBUs through GE P2P Ethernet optical access. The CBUs are then connected to wireless base stations through E1. The OLT connects wireless base stations to the core IP bearer network through optical access technologies. This implementation mode is not only simpler than traditional SDH/ATM private line technologies, but also drives down the costs of base station backhaul. FTTM is applicable to reconstruction and capacity expansion of mobile bearer networks. In this scenario, FTTM converges the fixed network and the mobile network on the bearer plane.

5.1.3 Configuring MDUs Cascaded to an OLT

MDUs are cascaded to an OLT through the P2P board, thereby saving upstream optical fibers and simplifying the network and service configuration.

Service Requirements

- MDU_1 and MDU_2 are connected to an OLT through GE cascading, implementing the Internet access service.
- The Internet access service is provided in the PPPoE dialing mode.

Figure 5-2 Network of MDUs cascaded to an OLT

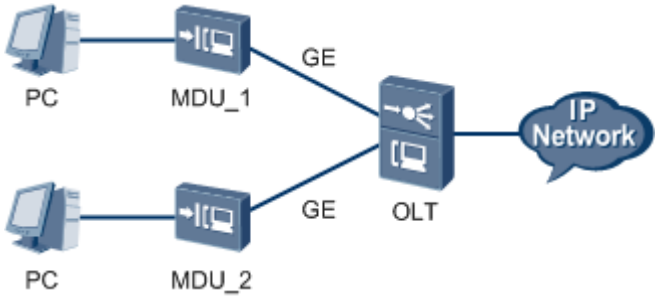


Table 5-1 Data plan

Item	Data
OLT	SVLAN ID: 100 SVLAN type: smart VLAN

 Copyright © Huawei Technologies Co., Ltd.

Item	Data
	CVLAN ID: 200
	Upstream port: 0/9/0
MDU_1	SVLAN ID: 200 SVLAN type: smart VLAN
	Upstream port: 0/0/1 **NOTE** The upstream ports vary with MDU type.
MDU_2	SVLAN ID: 200 SVLAN type: smart VLAN
	Upstream port: 0/0/1

Procedure

- Configure the OLT.
  a. Configure the port role.

     Configure the port role of the P2P board as a cascading port. The port roles of the P2P board are user port and cascading port. By default, the port role is user port.

     ```
     huawei(config)#interface eth 0/2
     huawei(config-if-eth-0/2)#network-role cascade
     huawei(config-if-eth-0/2)#quit
     ```

  b. Create a VLAN and add an upstream port to the VLAN.

     Create smart SVLAN 100. The upstream port is port 0/9/0.

     ```
     huawei(config)#vlan 100 smart
     huawei(config)#port vlan 100 0/9 0
     ```

  c. Configure a service port.

     Add the service port to the SVLAN by using default traffic profile 6. The CVLAN ID is 200, the same as the upstream VLAN ID of the MDU. MDU_1 and MDU_2 are connected to ports 0/2/0 and 0/2/1 of the OLT respectively.

     ```
     huawei(config)#service-port vlan 100 eth 0/2/0 multi-service user-vlan 200  rx-cttr 6 tx-cttr 6
     huawei(config)#service-port vlan 100 eth 0/2/1 multi-service user-vlan 200  rx-cttr 6 tx-cttr 6
     ```

  d. Save the data.

     ```
     huawei(config)#save
     ```

- Configure the MDUs.

  The configurations of MDU_1 and MDU_2 are the same. The configuration of MDU_1 is used as an example.

  a. Create a VLAN and add an upstream port to the VLAN.

     Create smart SVLAN 200. The upstream port is port 0/0/1.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

The SVLAN of the MDU must be the same as the CVLAN of the OLT.

```
huawei(config)#vlan 200 smart
huawei(config)#port vlan 200 0/0 1
```

b. Configure a service port.

According to actual conditions, an MDU supports multiple access modes. In this example, the ethernet port 0/3/1 is used. For other access modes, see the corresponding configuration guide of the MDU.

```
huawei(config)#service-port vlan 200 eth 0/3/1 multi-service user-vlan untagged  rx-cttr 6 tx-cttr 6
```

c. Save the data.

```
huawei(config)#save
```

----End

Result

On the PC, the Internet access service is provided in the PPPoE dialing mode.

Configuration File

Configure the OLT:

```
interface eth 0/2
network-role cascade
quit
vlan 100 smart
port vlan 100 0/9 0
service-port vlan 100 eth 0/2/0 multi-service user-vlan 200 rx-cttr 6 tx-cttr 6
service-port vlan 100 eth 0/2/1 multi-service user-vlan 200 rx-cttr 6 tx-cttr 6
save
```

Configure the MDU:

```
vlan 200 smart
port vlan 200 0/0 1
service-port vlan 200 eth 0/3/1 multi-service user-vlan untagged rx-cttr 6 tx-cttr 6
save
```

5.2 NE Cascading

NE cascading refers to a networking mode in which the MA5800 series are directly connected to each other through the ethernet port on the board. NE cascading saves the upstream optical fiber resources of the access node.

5.2.1 Introduction to NE Cascading

Definition

NE cascading refers to a networking mode in which the MA5800 series are directly connected to each other through the ethernet port on the board.

 Copyright © Huawei Technologies Co., Ltd.

Purpose

NE cascading makes the networking of the MA5800s more flexible, and saves the upstream optical fiber resources of the access node. In addition, remote subtending saves the convergence devices at the central office (CO), simplifies topology, and facilitates service configuration.

5.2.2 Principle of NE Cascading

Cascade network configurations make the networking of the MA5800s more flexible, saving the upstream optical fiber resources of the access point. The remote cascade network configurations save the aggregation equipment resource, simplify the networking, and facilities the service configuration.

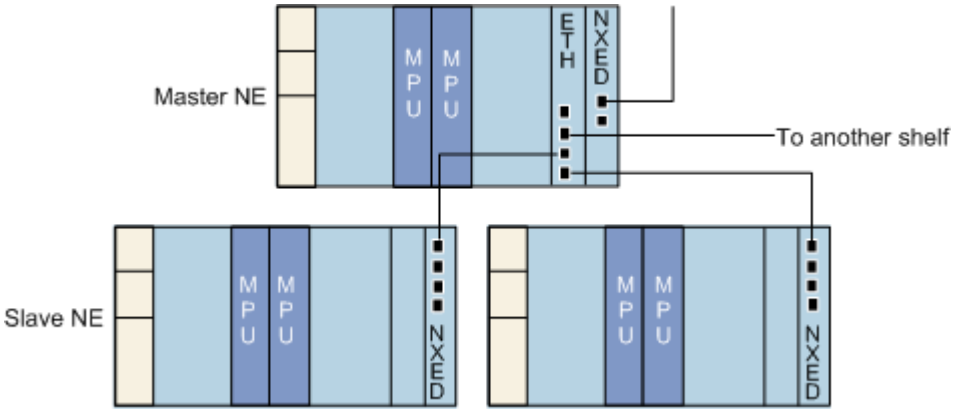
According to the location, the NE-level cascade network supported by the MA5800 consists of the **local cascading** and **remote cascading**.

- Local cascading refers to the cascading of multiple MA5800 subracks in a cabinet or the cascading of multiple local cabinets.
- Remote cascading refers to the cascading of which the remote MA5800 subrack is cascaded by using the optical fiber.

According to the connection mode, the NE-level cascade network supported by the MA5800 consists of the **star cascading** and **daisy chain cascading**.

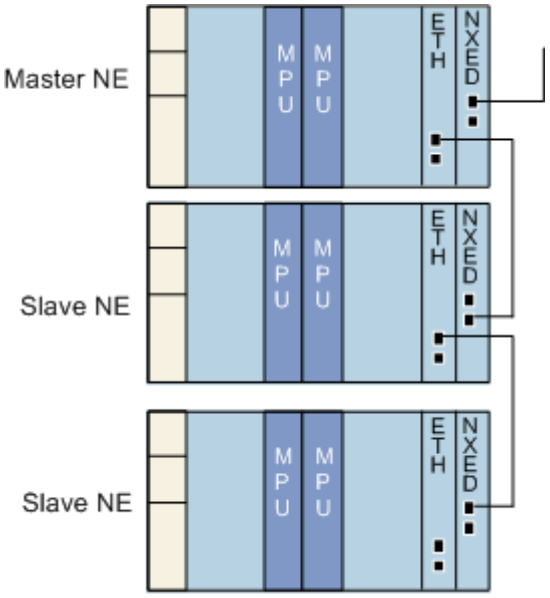
- In star cascading, multiple MA5800s form a star topology with the upstream device and connects to the upstream device through the MA5800. This networking features high reliability and simple network.
- In daisy chain cascading, multiple MA5800s form a chain topology with the upstream device and connects to the upstream device through one MA5800. This networking features low reliability, especially, if one device is faulty, all its downstream cascaded devices have upstream transmission failures.

Figure 5-3 Star cascading



Copyright © Huawei Technologies Co., Ltd.

Figure 5-4 Daisy chain cascading



5.2.3 Configuring NE Cascade and Uplink Transmission

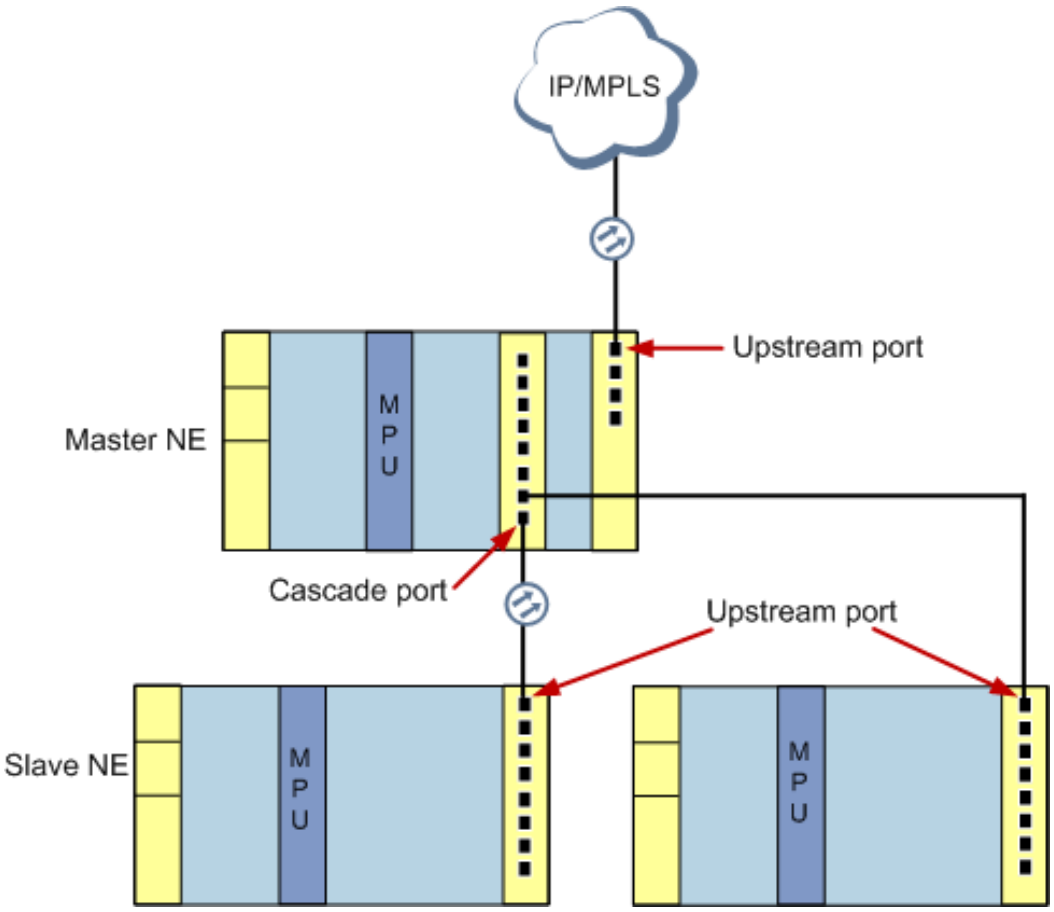
The MA5800s (NEs) can be directly connected to each other through the Ethernet port. Cascading saves the upstream optical fibers and simplifies networking and service configuration.

Context

The cascade ports on master NEs and upstream ports on slave NEs must have the same port type, port rate, and port duplex mode.

 Copyright © Huawei Technologies Co., Ltd.

Figure 5-5 Cascading network



NOTE

Configure the network role of a board first. Otherwise, the network role may fail to be switched if other configurations have been performed.

Procedure

- Configure the master NE
  a. Set the network role of the upstream board or port of the master NE.

     Enter the specific board mode according to the upstream board types. Run the **network-role [port-id] uplink** command to set the network role of the board or port to upstream.
  b. Set the network role of the cascade board or port of the master NE.

     Enter the specific board mode according to the cascade board types. Run the **network-role [port-id] cascade** command to set the network role of the board or port to cascading.
  c. Configure the VLAN of the master NE.

     The VLAN type is smart, and the VLAN attribute is common. For details about the configuration, see **Configuring a VLAN**.
  d. Add an upstream port to the VLAN of the master NE.

 Copyright © Huawei Technologies Co., Ltd.

According to the upstream board types, run the **port vlan** command in global config mode to add the upstream port to the VLAN.

e. Add a cascade port to the VLAN of the master NE.

According to the cascade board types, run the **port vlan** or **service-port** command in global config mode to add the cascade port to the VLAN.

NOTE

For boards OGHK and OXHD, run the **service-port** command to create service flows in Ethernet access.

- Configure the slave NE
  a. Set the network role of the upstream board or port of the slave NE.

Enter the specific board mode according to the upstream board types. Run the **network-role [port-id] uplink** command to set the network role of the board or port to upstream.

b. Configure the VLAN of the slave NE. The VLAN of the slave NE is the same as the VLAN of the master VLAN.

The VLAN type is smart, and the VLAN attribute is common. For details about the configuration, see **Configuring a VLAN**.

c. Add an upstream port to the VLAN of the slave NE.

According to the upstream board types, run the **port vlan** command in global config mode to add the upstream port to the VLAN.

----End

Example

Assume that master NE huawei_A and slave NE huawei_B are cascaded through the OXHD board. To add upstream port 0/3/1 and cascade port 0/2/0 of huawei_A to VLAN 100, and add upstream port 0/3/1 of huawei_B to VLAN 100, do as follows:

```
huawei_A(config)#interface eth 0/2
huawei_A(config-if-eth-0/2)#network-role cascade
huawei_A(config-if-eth-0/2)#quit
huawei_A(config)#vlan 100 smart
huawei_A(config)#port vlan 100 0/3 1
huawei_A(config)#service-port vlan 100 eth 0/2/0 multi-service user-vlan 100

huawei_B(config)#vlan 100 smart
huawei_B(config)#port vlan 100 0/3 1
```

5.3 Ethernet OAM

Ethernet Operation, Administration, and Maintenance (OAM) is used to operate, administrate, and maintain Ethernet networks. It covers three features: connectivity fault management (CFM), Ethernet in the first mile (EFM), and performance monitoring (PM).

 Copyright © Huawei Technologies Co., Ltd.

5.3.1 Introduction

Ethernet is a widely used local area network (LAN) technology featuring rich bandwidth and low cost. As the application of Ethernet is spanning from carrier networks to metropolitan area networks (MANs) and wide area networks (WANs), network administration and maintenance become increasingly important. However, traditional Ethernet is unable to provide end-to-end (E2E) service management, fault detection, or performance monitoring, which are instead implemented by IP-based mechanisms. These mechanisms (such as ping messages) at the IP layer are used to operate and manage the entire network, but they cannot meet operation and maintenance requirements of the Ethernet bottom layer that has no IP addresses. Against this backdrop, the IEEE and ITU-T have done a lot of researches and worked out their own Ethernet OAM standards.

Ethernet Operation, Administration, and Maintenance (OAM) covers three features: connectivity fault management (CFM), Ethernet in the first mile (EFM), and performance monitoring (PM). The following table describes the main functions of these features.

Feature	Description	Application Scenario
CFM	CFM is defined in the IEEE 802.1ag (802.1ag for short). It is an OAM feature for detecting and locating Ethernet connectivity faults by including the functions of continuity check (CC), loopback (LB), linktrace (LT), and so on. The ITU-T Y.1731 (Y.1731 for short) covers all the CFM functions defined in the 802.1ag and adds more information types to enhance the CFM capability.	CFM mainly applies to access and convergence nodes. It is used to monitor connectivity on the entire network and to locate connectivity faults for end-to-end (E2E) Ethernet links. For example, in IPTV services, CFM can be used to monitor network connectivity between a user terminal and the service provider's IPTV server and to determine the location of a fault.
EFM	EFM is defined in the IEEE 802.3ah (802.3ah for short). It is an OAM feature for detecting quality and connectivity of last-mile Ethernet links.	EFM packets cannot be forwarded across multiple hops. Therefore, EFM mainly applies to the user access network (that is, the last mile) and also to the Ethernet physical link that directly connects two devices. For example, EFM can be used to monitor the link between a user access board and a user terminal and to notify the system of link exceptions using Ethernet OAM link events. Link exceptions can be link interruptions or that the number of received error packets exceeds the preset threshold due to deteriorating link quality.

 Copyright © Huawei Technologies Co., Ltd.

Feature	Description	Application Scenario
PM	PM is defined in the Y.1731. It is an OAM feature for measuring and testing Ethernet performance, including loss measurement (LM), delay measurement (DM), and throughput measurement.	PM is mainly used to evaluate the network performance and quality and to discover latent network defects (such as an increasing packet loss rate due to line aging), reducing user complaints and improving quality of service (QoS).

NOTE

This chapter includes all the Ethernet OAM features that are supported by access devices. Some features and commands described in this chapter might be not supported by some specific devices. For details of which functions are supported by a specific device, refer to the specification list of the device.

5.3.2 Ethernet OAM Reference Standards and Protocols

Table 5-2 Reference standards and protocols of the Ethernet OAM feature

Ethernet OAM Type	Reference Standards and Protocols
CFM	• ITU-T Y.1731: OAM functions and mechanisms for Ethernet based networks • IEEE 802.1ag-2007 VLAN Amendment: 5 Connectivity Fault Management
EFM	IEEE 802.3ah: Operations, Administration, and Maintenance (OAM)
PM	ITU-T Y.1731: OAM functions and mechanisms for Ethernet based networks

NOTE

Connectivity fault management (CFM) is implemented according to the formal IEEE 802.1ag-2007 and is not compatible with Draft6.0.

ITU-T Y.1731 mentioned in this document is of version 07/2011.

5.3.3 Differences in Implementing Y.1731 and 802.1ag on Access Device

The ETH OAM of the access device is implemented based on concepts defined by 802.1ag. For concepts defined by both 802.1ag and Y.1731, there is no difference in implementation. For concepts defined by Y.1371 but not defined in 802.1ag, the similar concepts in 802.1ag are used. The following table describes the differences

 Copyright © Huawei Technologies Co., Ltd.

in implementing related concepts of Y.1731 and 802.1ag on the access device. The similar concepts of 802.1ag and Y.1731 are in the same row.

Concept of 802.1ag	Concept of Y.1731	Implementation of Differences Between Y.1731 and 802.1ag
Maintenance domain (MD)	No similar definition	For Y.1731, **MD name** is configured to **no name**.
Maintenance association (MA)	Maintenance entity group (MEG)	For Y.1731, ensure that the following requirements are met when configuring an MA. • **MA ID** is configured to **MEG ID**. • **MA name** must use the ICC-based format.
Maintenance point (MP)	MEG point (MP)	There is no difference in implementing MP in Y.1731 and 802.1ag.
Maintenance association end point (MEP)	MEG end point (MEP)	There is no difference in implementing MEP in Y.1731 and 802.1ag.
Maintenance association intermediate point (MIP)	MEG intermediate point (MIP)	There is no difference in implementing MIP in Y.1731 and 802.1ag.

5.3.4 CFM (802.1ag and Y.1731)

Connectivity fault management (CFM) is defined in the IEEE 802.1ag (802.1ag for short). It is an OAM feature for detecting and locating Ethernet connectivity faults. The ITU-T Y.1731 (Y.1731 for short) covers all the CFM functions defined in the 802.1ag and adds more information types to enhance the CFM capability.

5.3.4.1 CFM Introduction

The IEEE and ITU-T have worked out their own standards for monitoring Ethernet connectivity and identifying connectivity faults for E2E Ethernet links.

- The 802.1ag (also known as CFM) defines Ethernet Operations, Administration, and Maintenance (OAM) functions, which are used for detecting and identifying link connectivity faults on a network involving multiple independent maintenance organizations.
- The Y.1731 covers all functions defined in the 802.1ag and enhances some functions.

The access device supports CFM defined in the 802.1ag and Y.1731. **Table 5-3** describes functions of CFM.

 Copyright © Huawei Technologies Co., Ltd.

Table 5-3 Functions of CFM

Function	Description
Continuity check (CC)	Monitors connectivity of Ethernet links in real time and reports link exceptions by alarms.
LinkTrace (LT)	• Obtains MAC addresses of intermediate devices along the Ethernet link between two devices. • Determines the location of a link fault.
Loopback (LB)	Unicast LB • Checks connectivity of the Ethernet link between two devices. • Determines the location of a link fault. Multicast LB • Checks connectivity of Ethernet links between one device and multiple other devices. • Obtains the MAC address of a remote device that has the same maintenance level as the local device. Bidirectional diagnostics test: Checks the packet loss rate and bit error rate of the bandwidth-specific bidirectional Ethernet link between two devices to determine whether the bandwidth meets user requirements.
AIS (Alarm Indication Signal)	• Suppress alarms. • Send link fault information to the peer with higher-level.
LLF (Link Loss Forwarding)	Upon link fault detection, the device sends information to the peer end which is still communicated with the device. Then the peer end can react such as shutting down the port automatically. LLF is often used in the FTTO scenarios of national broadband network. When the leased channel of the retail service provider (RSP) is faulty, LLF function makes it possible that the RSP and its users are informed by the operator immediately.

Y.1731 and 802.1ag have the following differences in implementing the CFM function.

- The continuity check (CC) function and principle defined by Y.1731 and 802.1ag are the same only with one difference in the format of the continuity check message (CCM). Because of the difference, there are some limitations in usage of the CFM function when the access device is interconnected with the peer end that only supports Y.1731.
  - When MD is being configured, configure **MD name** to no name.
  - When MA is being configured, MA ID is identical to MEG ID. **MA name** must use the ICC-based format.

 Copyright © Huawei Technologies Co., Ltd.

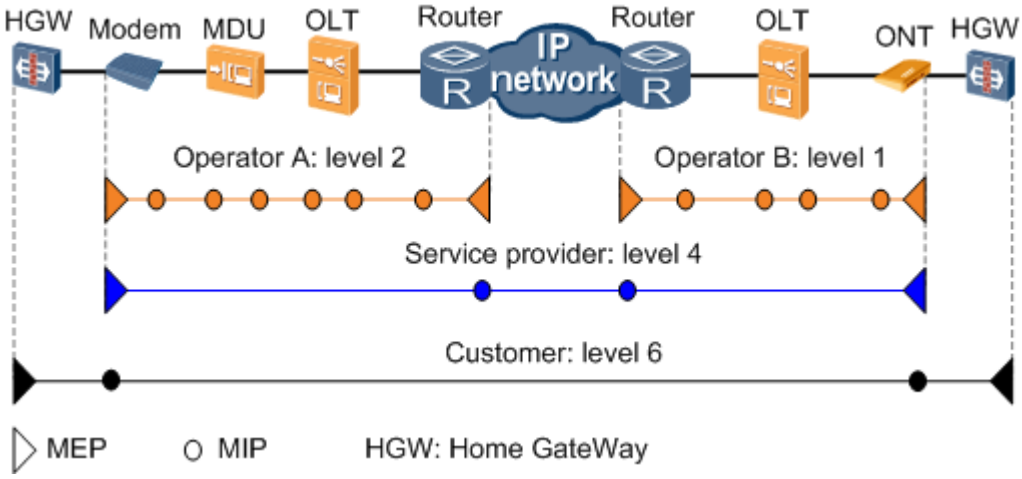
- 802.1ag defines the unicast LB. Y.1731 defines the unicast LB, multicast LB, and bidirectional diagnostics test. In addition, Y.1731 provides the extended loopback message (LBM), and the size and sending interval of LBMs can be set and disorder in the loopback reply (LBR) message returned by the peer end can be detected.
  - No matter whether the peer end device supports Y.1731 or 802.1ag, when implementing an LB (including the unicast LB, multicast LB, and bidirectional diagnostics test), the access device can receive the LBR message returned by its peer end. In addition, the test result does not vary with the protocol used at the peer end.
  - No matter whether the peer end device supports Y.1731 or 802.1ag, the access device can correctly reply with the LBR messages.
- Y.1731 and 802.1ag have no difference in implementing the LT function. No matter whether the peer end device supports Y.1731 or 802.1ag, the access device can implement the link trace (LT) and correctly reply with the linktrace reply (LTR) messages.
- ETH-AIS is only defined in Y.1731.
- LLF is a function defined by Huawei. It is implemented based on the information format defined in 802.1ag.

5.3.4.2 CFM Network Application

In carrier-class Ethernet networks, different organizations are required to provide different administration and maintenance scopes and contents. Therefore, carrier-class Ethernet needs to be maintained hierarchically. Three types of organizations are generally involved in carrier-class Ethernet services: customers (for example, private network users), service providers, and network operators. Customers purchase Ethernet services from service providers, service providers can use their own networks or leased networks to provide end-to-end (E2E) Ethernet services, while network operators provide networks for transmitting services.

The following picture shows a typical CFM network. In the figure, locations of maintenance association end points (MEPs) and maintenance domain intermediate points (MIPs) are marked, and different maintenance domains (MDs) are configured for three maintenance levels (customer, service provider, and network operator).

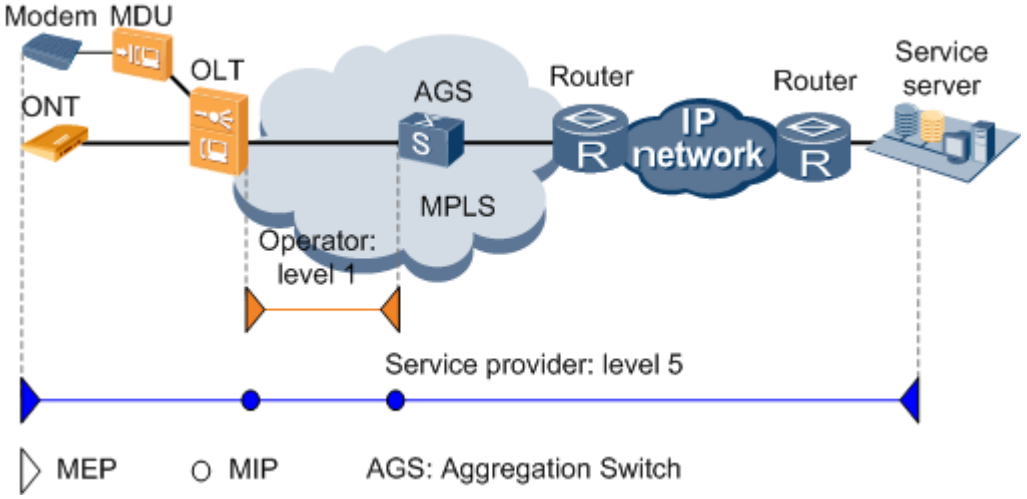
Figure 5-6 Typical CFM network



 Copyright © Huawei Technologies Co., Ltd.

CFM also applies to MPLS networks for connectivity diagnosis, as shown in the picture below.

Figure 5-7 Typical CFM network



5.3.4.3 CFM Basic Concepts

This topic describes some basic concepts of the connectivity fault management (CFM), such as the maintenance domain (MD), maintenance association (MA), maintenance association end point (MEP) and maintenance association intermediate point (MIP) that we need to understand before getting to know the working principle of the CFM function.

MD

A maintenance domain (MD) is used to divide the network covered by the connectivity fault test.

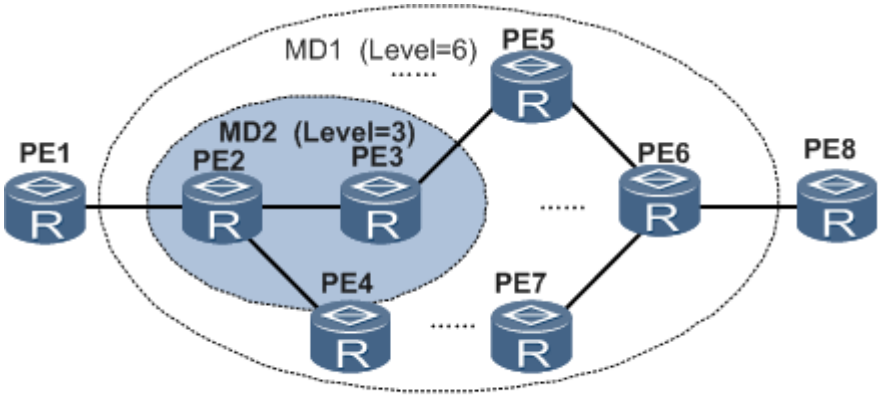
In order to locate a fault precisely, the concept of **level** is introduced in MD. An MD provides eight levels, represented by the integer number 0-7. A larger number indicates a higher level and a larger-range MD. An MD can be embedded into or neighbor on another MD but an MD cannot cross another. A low-level MD is embedded into a high-level MD, but a high-level MD cannot be embedded into a low-level MD.

The MD level makes the fault locating precise and easy. As shown in **Figure 5-8**, MD2 is embedded into MD1. If MD1 detects a fault, it is suspected that the links or the devices from PE2 to PE6 are faulty. If MD2 does not detect any fault, PE2, PE3 and PE4 work properly. Therefore, the fault is narrowed down to the links between PE5, PE6, and PE7 or these devices.

In actual application, if an MD contains another small-range MD and the connectivity test is performed for the large-range MD, CFM packets need to traverse the small-range MD. To achieve that purpose, set the level of the large-range MD higher than that of the small-range MD. For example, in the network as shown in **Figure 5-8**, MD1 contains MD2 and CFM packets of MD1 need to traverse MD2. The level of MD1 is set to 6 and the level of MD2 is set to 3. In this way, CFM packets of MD1 can traverse MD2 and CFM for the entire MD1 can be implemented. Moreover, CFM packets of MD2 will not spread to MD1.

Copyright © Huawei Technologies Co., Ltd.

Figure 5-8 MD levels



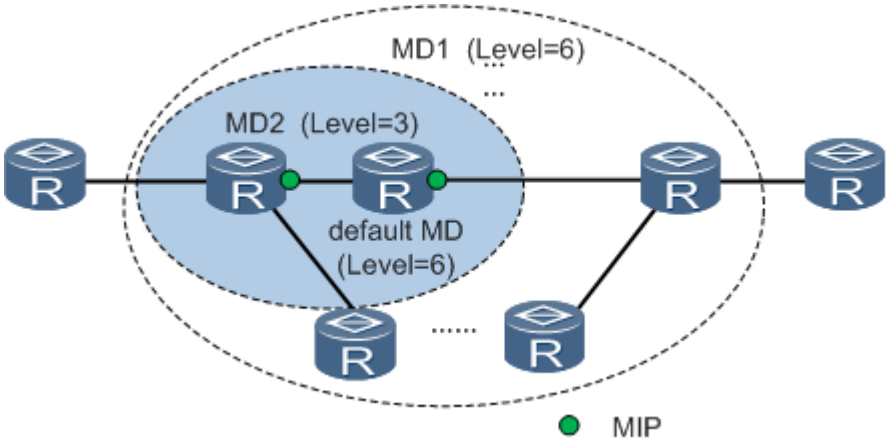
CFM packet interaction and basic CFM functions are implemented based on MDs. Proper MD planning can help network administrators locate faults quickly.

Default MD

The default MD is used for the high-level MD to sense the internal topology of the low-level MD.

As shown in **Figure 5-9**, a low-level MD is embedded into a high-level MD, and devices in the high-level MD may be the edge and intermediate devices in the low-level MD. When CFM protocol packets of the high-level MD traverse the low-level MD, the packets are transmitted transparently. If the default MD is not configured and the high-level MD needs to sense the internal topology of the low-level MD, a maintenance association intermediate point (MIP) with a specific level needs to be created on the specific port of the device in the low-level MD and the MIP is used to reply to devices in the high-level MD with loopback reply (LBR) or linktrace reply (LTR) messages.

Figure 5-9 Default MD



If the default MD with the same level as the high-level MD is configured on the devices in the low-level MD, MIPs of corresponding levels based on default MDs are generated to reply to devices in the high-level MD with LBR or LTR messages.

 Copyright © Huawei Technologies Co., Ltd.

In this way, the high-level MD can sense the topology changes in the low-level MD and CFM can be implemented in the entire MD1.

The level of the default MD is the same as that of the high-level MD but is higher than the levels of all MDs to which the MEPs configured on the device belong.

MA

An MA specifies the range a CCM can reach.

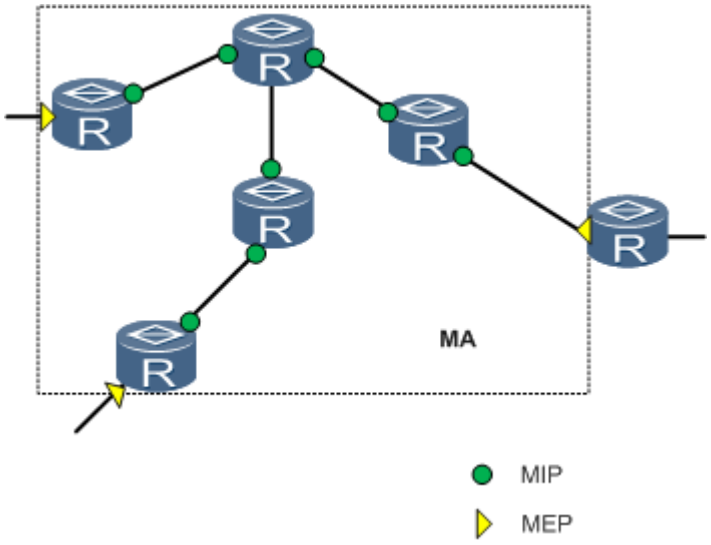
One or more MAs can be configured in an MD according to the actual needs. Each MA is a grouping of some maintenance points (MPs) in an MD. The level of an MA is the same as the level of the MD to which it belongs.

An MA serves a certain service (such as a VLAN): All packets sent by MPs in an MA carry the service tag, and an MP can receive packets sent by other MPs in the same MA.

MP

MPs are configured on ports and belong to a certain MA. There are two MPs: MEP and MIP. **Figure 5-10** shows the MPs.

Figure 5-10 MP



MEP

MEPs are end points of the MD and MA, and they specify the range and boundary of the MD and MA.

The level of an MEP is identical to the level of the MD to which it belongs. The level of an MEP determines the level of packets it can process.

- When an MEP receives a packet whose level is higher, it does not process this packet and forwards the packet along its original path so that the packet can traverse the MD.
- If an MEP receives a packet whose level is the same or lower, the MEP will process the packet.

 Copyright © Huawei Technologies Co., Ltd.

The levels of packets sent by an MEP are the same as the level of the MEP.

MEPs are located on ports of devices and need to be created by users manually.

The MEP on any device that runs Ethernet CFM is called the local MEP. The MEPs on the other devices in the same MA are the remote maintenance association end points (RMEPs) to the local MEP.

MIP

MIPs are inside an MA. Network management can be improved by deploying multiple MIPs among MEPs. More MIPs bring about stronger control and improved management over networks. For profitable services that key accounts care about, carriers deploy much more MIPs.

MIPs are located on ports of devices, and they are generated automatically according to specific rules but cannot be created manually.

5.3.4.4 CFM Principles

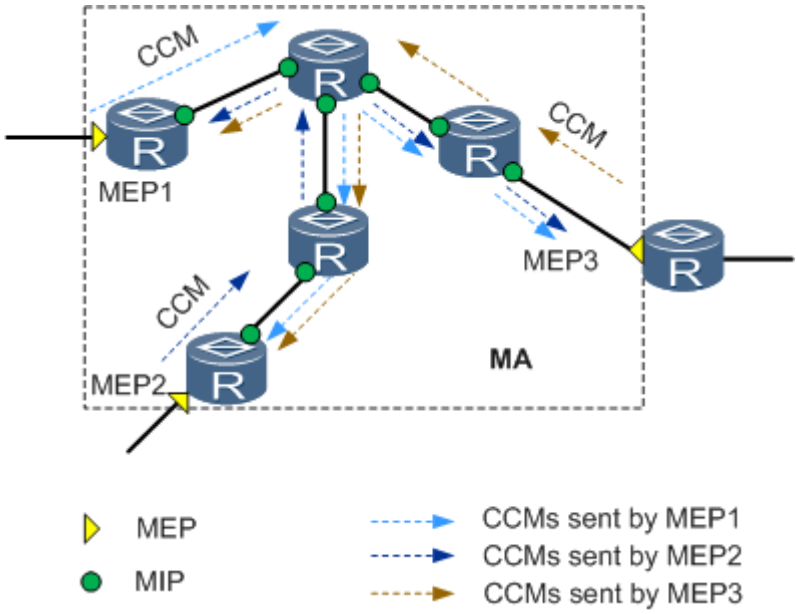
This topic describes the basic concepts and principles of CFM based on 802.1ag. Because the connectivity fault management (CFM) function defined by Y.1731 is similar to the one defined by 802.1ag, you can refer to the similar concepts of 802.1ag to get to know the mechanism of Y.1731 unless otherwise specified.

5.3.4.4.1 CC Principle

Continuity check (CC) is used to monitor connectivity of Ethernet links in real time and to report link exceptions by alarms.

A maintenance association end point (MEP) periodically sends multicast continuity check messages (CCMs) to other MEPs in the same maintenance association (MA). If an MEP fails to receive expected CCMs or receives unexpected CCMs within a certain period, the MEP considers that it has detected a connectivity fault in its MA. **Figure 5-11** illustrates the CC principle.

Figure 5-11 CC principle



 Copyright © Huawei Technologies Co., Ltd.

1. MEP database setup

   Each device with Ethernet connectivity fault management (CFM) enabled has an MEP database. The database records the MEPs (local MEPs) configured on the local device and other devices' MEPs (that is, RMEPs) in the same MA. Local MEPs and RMEPs are manually configured but automatically recorded to the MEP database.

2. CCM generation

   An MEP generates and sends CCMs. As shown in **Figure 5-11**, MEP1, MEP2, and MEP3 are in the same MA. After the function of sending CCMs is enabled, MEP1, MEP2, and MEP3 periodically multicast CCMs to each other at the same interval.

   All maintenance domain intermediate points (MIPs) and MEPs in an MA receive CCMs, with no need to send a reply.

   A CCM carries its maintenance level information. The level of a CCM is equal to the level of the MEP sending the CCM.

3. CCM termination

   An MEP terminates CCMs. If an MEP receives a CCM whose level is higher than the MEP's, the MEP forwards this CCM. If an MEP receives a CCM whose level is lower than or equal to the MEP's, the MEP does not forward this CCM, ensuring that CCMs in a lower-level maintenance domain (MD) are not sent to a higher-level MD.

4. Fault identification

   If an MEP fails to receive expected CCMs or receives unexpected CCMs within a certain period, the MEP considers that it has detected a connectivity fault in its MA and reports a link fault alarm. When multiple MEPs in an MD send CCMs, multipoint-to-multipoint (MP2MP) link connectivity is checked.

NOTE

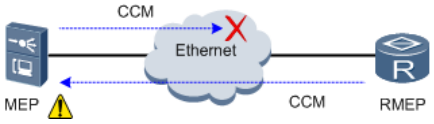
CC can determine whether a fault occurs but cannot identify the fault location. In this case, loopback (LB) and linktrace (LT) can be used to determine the fault location.

Table 5-4 describes the alarms may be generated when link faults occur on a network.

Table 5-4 CC Alarm

Alarm Name	Principle	Troubleshooting
0x2f210001 Loss of Ethernet connectivity check message	If an MEP does not receive a CCM from an RMEP within a period equal to 3.5 times the CC interval, the MEP considers that the RMEP is lost, and the device on which the MEP is configured generates an alarm. The following figure shows how the alarm is generated. 	When the alarm is generated, the link from the RMEP to the MEP is faulty. In this case, LB or LT can be used on the MEP to determine the fault location.

 Copyright © Huawei Technologies Co., Ltd.

Alarm Name	Principle	Troubleshooting
0x2f210004 Reception of Ethernet connectivity check message with the RDI bit set	When an MEP receives CCMs from its RMEP but the RMEP fails to receive CCMs from the MEP within a period equal to 3.5 times the CC interval, the RMEP considers that the MEP is lost and sets the RDI bit in its CCMs to be sent to 1. If the MEP receives a CCM with the RDI bit set to 1, the device on which the MEP is configured generates an alarm. The following figure shows how the alarm is generated.  	When the alarm is generated, the link from the MEP to the RMEP is faulty but the link from the RMEP to the MEP is normal. That is, a unidirectional connectivity fault occurs on the link. In this case, LB or LT can be used on the MEP to determine the fault location.
0x2f210002 Reception of invalid Ethernet connectivity check message	When an MEP receives CCMs with errors from its RMEP, the device on which the MEP is configured generates an alarm indicating reception of incorrect Ethernet OAM CCMs. Error types include: • Unexpected Periodicity defect (dUNP): The interval of sending CCMs is different from the CC interval of the MEP that receives the CCMs. • Unexpected MEP defect (dUNM): The MEP ID of CCMs is not included in the RMEP ID list of the MEP that receives the CCMs.	When the alarm is generated, the link between the MEP and the RMEP is not interrupted but the MEP receives incorrect CCMs. In this case, the network CFM data plan needs to be checked.
0x2f210003 Reception of Ethernet cross connect connectivity check message	When a CCM received by an MEP from its RMEP contains an item (MD name type, MD name, MA name type, or MA name) different from that configured on the MEP, the device on which the MEP is configured generates an alarm indicating reception of cross connect Ethernet OAM CCMs. The possible causes of the alarm are as follows: • The CCM with the level lowered than the local MEP is received. • When the CCM with the same level is received, the name or type of the peer MA or MD does not match the name of the local end.	When the alarm is generated, the link between the MEP and the RMEP is not interrupted but the MEP receives incorrect CCMs. In this case, the CFM data plan needs to be checked on the MEP and the RMEP.

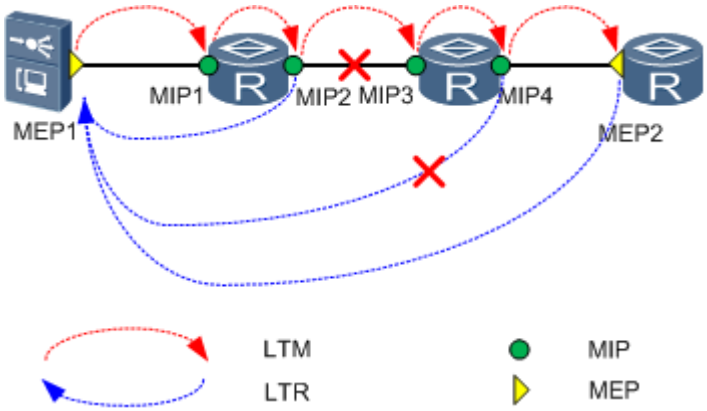
Copyright © Huawei Technologies Co., Ltd.

5.3.4.4.2 LT Principle

Linktrace (LT) is used to obtain MAC addresses of intermediate devices along the Ethernet link between two devices and also determine the location of a link fault.

Along the link between the maintenance association end point (MEP) LT initiator and the target MEP, every device that generates a maintenance domain intermediate point (MIP) sends a linktrace reply (LTR) to the initiator and forwards the linktrace message (LTM) until the LTM reaches its target. The MEP LT initiator, through LTRs, obtains MAC addresses and locations of all in-between devices as well as the link section where a fault occurs. **Figure 5-12** illustrates the LT principle.

Figure 5-12 LT principle



1. When all links are functioning properly, MEP1 sends an LTM to MEP2.
2. After all devices along the link between MEP1 and MEP2 receive the LTM, egress points (MIP2 and MIP4) return an LTR to MEP1, and forward the LTM with the time to live (TTL) field decremented by 1 to the next hop. An LTR contains the MAC address of the device sending the LTR.
3. After receiving the LTM, MEP2 stops forwarding the LTM but directly returns an LTR to MEP1.
4. As shown in **Figure 5-12**, when the link between MIP2 and MIP3 is faulty, MEP1 can receive an LTR from MIP2 but not from MIP4 after sending an LTM destined for MEP2. In this way, the fault location is determined.

5.3.4.4.3 LB Principle

Loopback (LB) is used to check connectivity of end-to-end (E2E) Ethernet links between devices and also determine the location of a link fault.

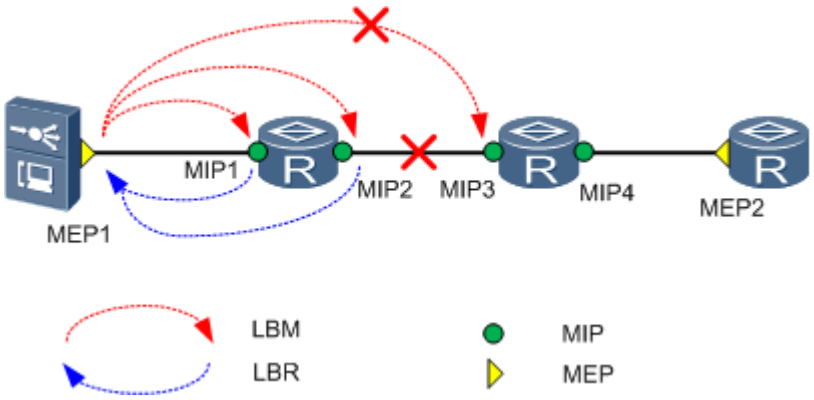
Unicast LB

Unicast LB, similar to the ping operation at the IP layer, is mainly used to check the status of connection between a local device and a remote device.

Figure 5-13 illustrates the principle of unicast LB.

Copyright © Huawei Technologies Co., Ltd.

Figure 5-13 Unicast LB principle



1. MEP1 sends a loopback message (LBM) to MIP1.
2. MEP1 receives a loopback reply (LBR) from MIP1.

   NOTE

   A maintenance domain intermediate point (MIP) only responds to the received LBM but does not forward the LBM to the next-hop MIP or the target maintenance association end point (MEP).
3. MEP1 sends an LBM to the next hop of MIP1, that is, MIP2.
4. MEP1 receives an LBR from MIP2.
5. MEP1 sends an LBM to the next hop of MIP2, that is, MIP3.
6. MEP1 fails to receive an LBR from MIP3 because the link between MIP2 and MIP3 is faulty.
7. MEP1 determines that the link between MIP2 and MIP3 is faulty.

In unicast LB, the destination MAC address of an LBM is a unicast MAC address. Therefore, before using unicast LB, ensure that:

- The MAC address or MEP index of MEP2 and the MAC addresses of all MIPs along the link between MEP1 and MEP2 are obtained if the connection status of all links between MEP1 and MEP2 is required for accurately determining the fault location. LT can be performed on MEP1 to obtain the MAC addresses of all MIPs along the link between MEP1 and MEP2.
- The MAC address or MEP index of MEP2 is obtained if only the connection status between MEP1 and MEP2 is required.

MEP1 can send multiple LBMs and, based on the number of received LBRs, determine whether packet loss occurs on the network.

Multicast LB

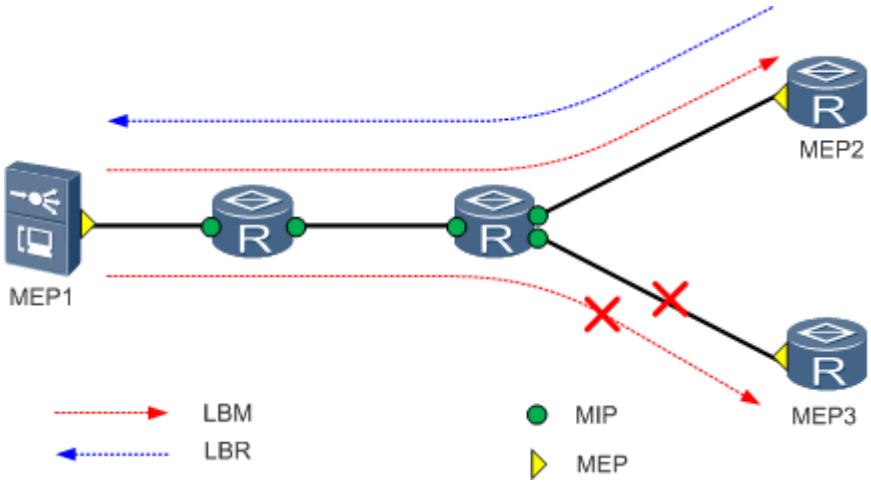
Multicast LB is used to check connectivity between an MEP and its peer MEPs and to obtain MAC addresses of those MEPs.

Multicast LB generally applies to scenarios in which multiple MEPs are deployed. In multicast LB, the MAC address of an LBM is a multicast MAC address in format of 01-80-C2-00-00-3x, where x is the level of the MEP initiating multicast LB. When the level of a peer MEP is equal to the level of the received multicast LBM,

 Copyright © Huawei Technologies Co., Ltd.

the peer MEP returns an LBR to the MEP initiating multicast LB. **Figure 5-14** illustrates the principle of multicast LB.

Figure 5-14 Multicast LB principle



1. MEP1 multicast an LBM.
2. MEP2 returns an LBR after receiving the LBM.
3. MEP3 fails to receive the LBM and therefore does not return an LBR because the link between MEP3 and MEP1 is interrupted.
4. MEP1, according to the received LBR, determines that the link between MEP1 and MEP2 is functioning properly but the link between MEP1 and MEP3 is faulty.

An LBR contains the MAC address of the remote MEP (RMEP) sending this LBR; therefore, multicast LB ca n also be used to obtain the MAC address of an RMEP.

As defined in the Y.1731, only one multicast LBM can be sent at a time. However, for compatibility with LB defined in the 802.1ag, the system supports a configurable number of multicast LBMs to be sent. An RMEP has a delay in responding to the received LBM. Therefore, an MEP may receive disordered LBRs in response to multiple LBMs it sends. That is, the serial numbers (SNs) of LBRs returned by an RMEP may not be consecutive. Therefore, it is recommended that the number of LBMs that can be sent at a time be set to 1 before multicast LB is performed.

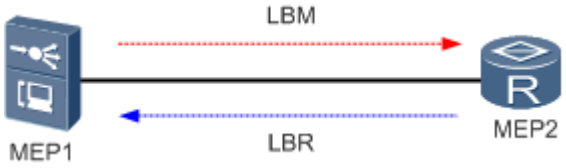
Bidirectional diagnostics test

Bidirectional diagnostics test is used to check the packet loss rate and bit error rate of a bandwidth-specific bidirectional link between two MEPs to determine whether the bandwidth meets user requirements.

Figure 5-15 illustrates the principle of bidirectional diagnostics test.

 Copyright © Huawei Technologies Co., Ltd.

Figure 5-15 Bidirectional diagnostics test principle



In bidirectional diagnostics test, an LBM and an LBR both carry a testing code, which is used for pseudo random binary sequence (PRBS) check. The following describes the process of bidirectional diagnostics test:

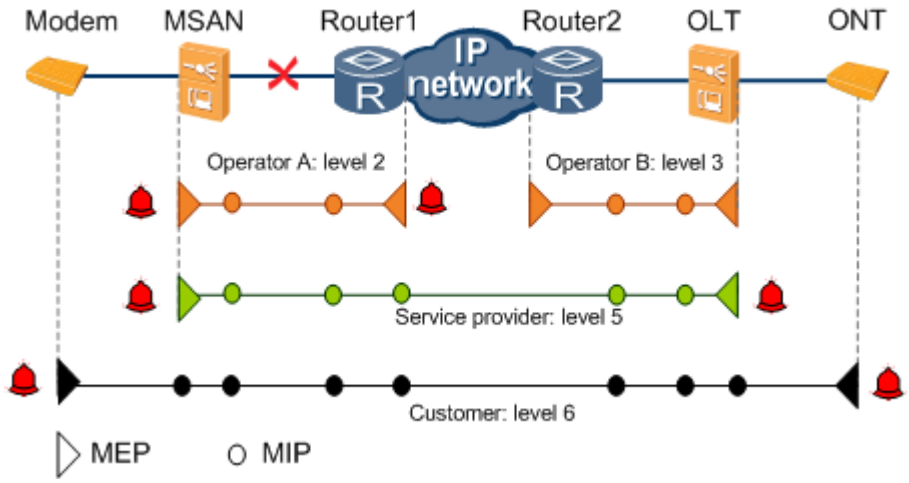
1. MEP1 sends a unicast LBM to MEP2.
2. MEP2 modifies the received LBM to an LBR and returns it to MEP1.
3. MEP1 analyzes the received LBR and determines whether packet loss and bit errors occur over the bidirectional link between MEP1 and MEP2.

5.3.4.4.4 AIS Principles

Alarm indication signal (AIS) is a dedicated function of Y.1731 and is not supported by 802.1ag. AIS is used to send link fault notifications to the client. AIS also suppresses redundant alarms.

As shown in **Figure 5-16**, when lower-level services encounter continuity check (CC) faults, maintenance association end points (MEPs) at each level detect CC faults through ETH CC and report fault alarms. Alarms generated in this scenario are mostly redundant alarms. Actually, the OAM engineer can locate the section of the link where the fault has occurred according to the alarm reported by the lowest-level MEP. For example, if the network of Operator A encounters a fault, the fault can be identified within this network, and there is no need to locate the fault in maintenance domains (MDs) owned by the service provider and customer.

Figure 5-16 Before AIS Is Enabled

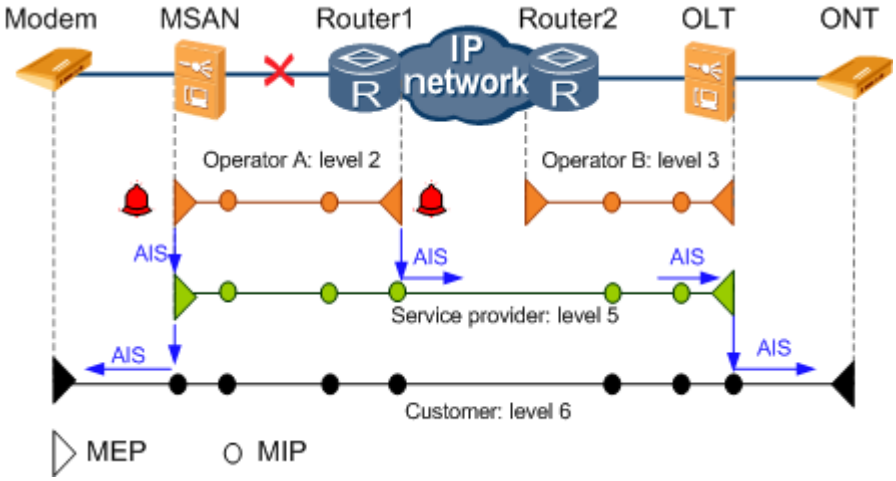


Alarm suppression can be performed through AIS. Specifically, after a lower-level MEP detects a CC fault, this MEP periodically sends AIS packets to the client MEG level in the reverse direction. After receiving these AIS packets, the client-level

 Copyright © Huawei Technologies Co., Ltd.

MEP no longer reports an alarm when detecting a CC fault. For example, MDs owned by the service provider and customer do not report any alarms.

Figure 5-17 After AIS Is Enabled



Fault types triggering AIS alarms include connection loss, crossed connection, errored MEP ID, errored period, and MAC status fault.

5.3.4.4.5 LLF Principles

Link loss forwarding (LLF) is a Huawei-defined Ethernet OAM function. It transmits fault information about Ethernet links according to message types defined in the 802.1ag protocol.

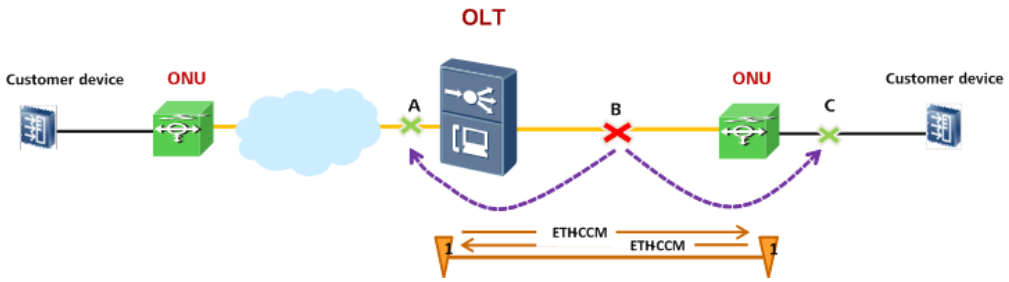
Usually, LLF is used in open access networks for national broadband programs. In fiber to the office (FTTO) scenarios, retail service providers (RSPs) set up networks of their own and rent access pipelines and device ports from operators. If pipelines rent to RSPs encounter faults, LLF can be used to notify RSPs and the RSPs' affected enterprise users so that they can know about the faults and take corresponding measures more responsively.

LLF sends the interface status TLV (defined in 802.1ag) carried in the continuity check messages (CCMs) to the peer maintenance end point (MEP). Then, the peer MEP judges whether to generate a fault alarm or fault clear alarm according to the received interface status TLV.

1. Maintenance scenario 1: The network is disconnected and the network-side port and user-side port are disabled. The network continuity is checked using the continuity check (CC) function of the MEP. Specifically, if the MEP on port A or port C detects a CC failure, port A or port C will be disabled; if the MEP on port A or port C detects a CC failure recovery, port A or port C will be enabled. **Figure 5-18** shows this scenario.

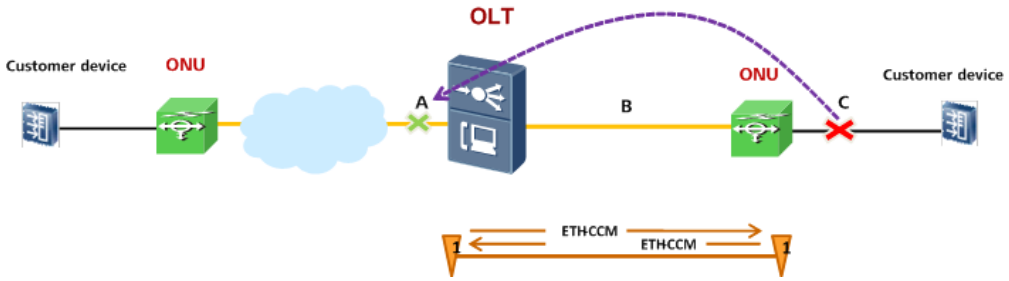
Copyright © Huawei Technologies Co., Ltd.

Figure 5-18 Disabling the network-side port and user-side port when the network is disconnected



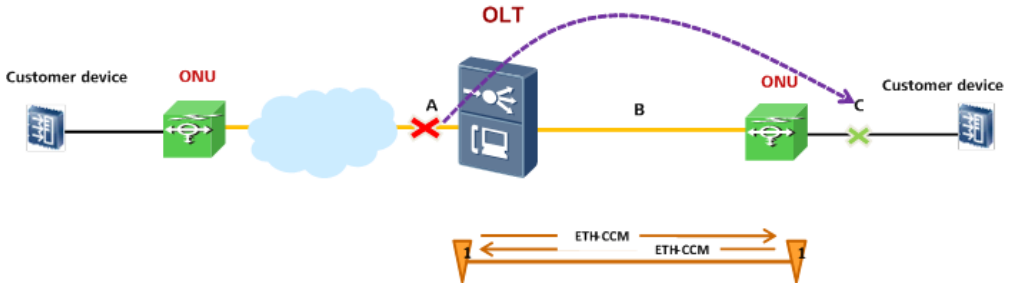
2. Maintenance scenario 2: The user-side port is disconnected and the corresponding network-side port is disabled. Specifically, when the user-side port C is disconnected, interface status TLV carried in the CCM is used to indicate whether the local port is faulty. If the MEP on port A receives a CCM carrying interface status TLV 2, the MEP will be notified that the user-side port C is faulty and will disable port A; if the MEP on port A receives a CCM carrying interface status TLV 1, the MEP will be notified that the user-side port C has recovered and will enable port A. **Figure 5-19** shows this scenario.

Figure 5-19 Disabling the corresponding network-side port when the user-side port is disconnected



3. Maintenance scenario 3: The network-side port is disconnected and the corresponding user-side port is disabled. Specifically, when the network-side port A is disconnected, interface status TLV carried in the CCM is used to indicate whether the local port is faulty. If the MEP on port C receives a CCM carrying interface status TLV 2, the MEP will be notified that the network-side port A is faulty and will disable port C; if the MEP on port C receives a CCM carrying interface status TLV 1, the MEP will be notified that the network-side port A has recovered and will enable port C. **Figure 5-20** shows this scenario.

Figure 5-20 Disabling the corresponding user-side port when the network-side port is disconnected

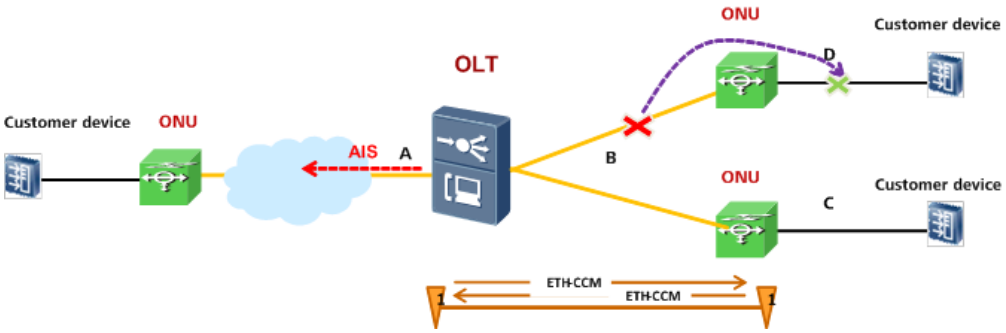


4. Maintenance scenario 4: In the aggregation scenario, the network is disconnected and the alarm indication signal (AIS) is sent. Specifically, in the

 Copyright © Huawei Technologies Co., Ltd.

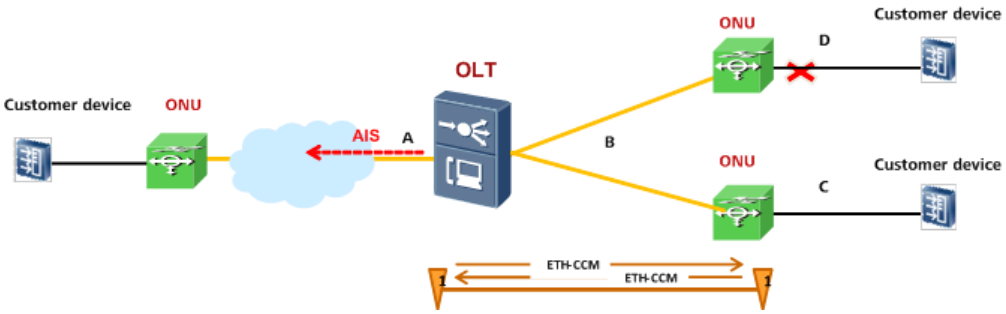
aggregation scenario, when services of a user in the network are interrupted, the network-side port cannot be disabled because services of the other users are still normal. In this case, the OLT sends the AIS to the normal ONU to notify a link abnormality. If link B is disconnected, the MEP on port A detects a failure and notifies the ONU of this failure; if the MEP on port D detects a CC failure, port D is disabled. After the fault is rectified, if the MEP on port D detects a CC failure recovery, port D will be enabled; if the MEP on port A detects a CC failure recovery, it stops sending AISs. **Figure 5-21** shows this scenario.

Figure 5-21 Sending AISs when the network is disconnected in the aggregation scenario



5. Maintenance scenario 5: In the aggregation scenario, the user-side port is disconnected and the AIS is sent. Specifically, in the aggregation scenario, when the user-side link is disconnected, the network-side port cannot be disabled because the other user-side links are still normal. In this case, the AIS is sent to notify the ONU of a link abnormality. When user-side port D is disconnected, interface status TLV carried in the CCM is used to indicate whether the local port is faulty. If the MEP on port A receives a CCM carrying interface status TLV 2, the MEP will be notified that the user-side port D is faulty and will send the AIS to notify the ONU of a failure; if the MEP on port A receives a CCM carrying interface status TLV 1, the MEP will be notified that the user-side port D has recovered and will stop sending the AIS. **Figure 5-22** shows this scenario.

Figure 5-22 Sending AISs when the user-side port is disconnected in the aggregation scenario

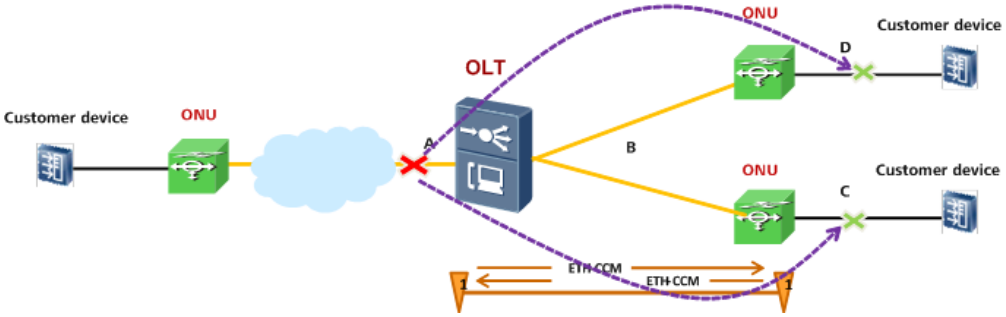


6. Maintenance scenario 6: In the aggregation scenario, the network-side port is disconnected and all the user-side ports are disabled. Specifically, in the aggregation scenario, when the network-side port A is disconnected, interface status TLV carried in the CCM is used to indicate whether the local port is faulty. If the MEP on port C or port D receives a CCM carrying interface status

 Copyright © Huawei Technologies Co., Ltd.

TLV 2, the MEP will be notified that the port C or port D is faulty and will disable port C or port D; If the MEP on port C or port D receives a CCM carrying interface status TLV 1, the MEP will be notified that port C or port D has recovered and will enable port C or port D. **Figure 5-23** shows this scenario.

Figure 5-23 Disabling all the user-side ports when the network-side port is disconnected in the aggregation scenario



The preceding scenarios which are achieved under the guideline of the following LLF principles:

1. The local MEP adds interface status TLV to the CCM according to the port link status and then sends the CCM.
2. The peer MEP determines to generate or not to generate a someMACstatusDefect alarm according to the interface status TLV carried in the CCM.

   The someMACstatusDefect alarm is generated by complying with the 802.1ag protocol. Specifically, when the port on a remote MEP is not UP, the MEP generates the someMACstatusDefect alarm. Alternatively, when none of the ports on the remote MEPs are in the forwarding state, the MEPs generate the someMACstatusDefect alarm.

Defect		Priority	
Variable	highestDefect (20.33.9)	highestDefectPri (20.33.8)	Importance
xconCCMdefect (20.23.3)	DefXconCCM	5	Most
errorCCMdefect (20.21.3)	DefErrorCCM	4	
someRMEPCCMdefect (20.33.5)	DefRemoteCCM	3	
someMACstatusDefect (20.33.6)	DefMACstatus	2	
someRDIdefect (20.33.7)	DefRDICCM	1	Least

3. When an MEP detects a CC alarm that has a higher severity than the associated alarm severity, the port connected to the MEP is automatically shut down. When the alarm is cleared, the port status recovers. A port can also be

 Copyright © Huawei Technologies Co., Ltd.

shut down through the CLI. Therefore, before performing configuration, know the relationships between the associated port shutdown function and the CLI-configured port shutdown function, as listed in the following table.

Associated Status	CLI-configured Status	Final Status
Shutdown	Shutdown	Shutdown
Shutdown	Undo shutdown	Shutdown
Undo shutdown	Shutdown	Shutdown
Undo shutdown	Undo shutdown	Undo shutdown
Associated port shutdown function disabled	Shutdown	Shutdown
Associated port shutdown function disabled	Undo shutdown	Undo shutdown

NOTE

Triggering the associated port shutdown function means to associate the MEP CC alarm status with the activation status of the port where the MEP resides. In other words, the port is automatically disabled or enabled according to the CC status. When association is enabled, the lowest CC alarm severity (2 by default) needs to be configured. Association is triggered when the severity of the CC alarm generated by the MEP is higher than or equal to the configured association alarm severity. When association is disabled, the port status is irrelevant to the CC status.

5.3.4.5 Configuring the Ethernet CFM OAM

CFM OAM is an OAM at the network level. It supports connectivity check, loopback testing and link tracking, and applies to the end-to-end fault detection in large-scale network.

Prerequisites

- ONT or Modem must support the 802.1ag protocol.
- Service configurations (VLAN configuration and service port configuration, for example) are finished, and the customer services are normal.

Networking

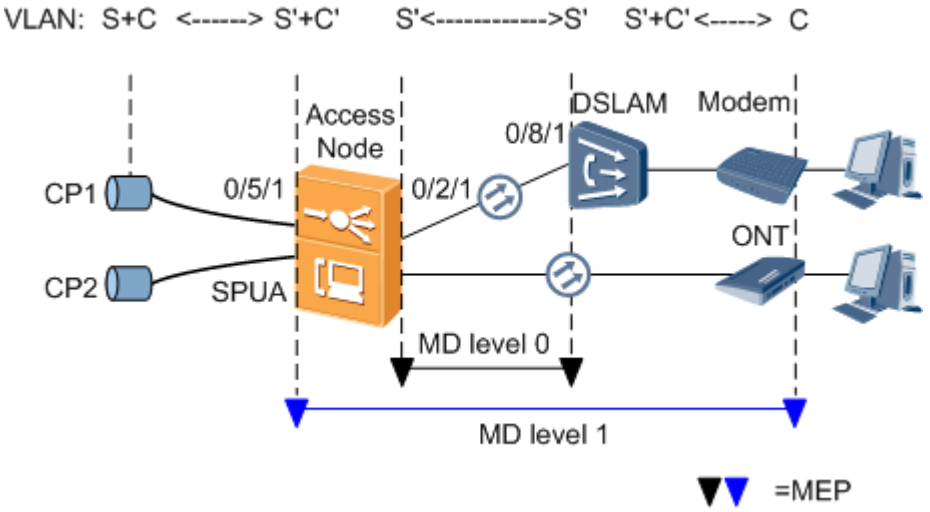
Figure 5-24 shows a typical FTTx integrated networking. Through the deployment of CFM OAM, service providers can detect connectivity between any two devices based on their needs. When there is a connectivity problem, the system generates an alarm reporting the fault location. Assume that the access node is an OLT.

- The OLT cascades to the DSLAM through a cascade board. An MD is deployed for the management channel between the cascading port (0/2/1) of OLT and the upstream port (0/8/1) of DSLAM.

 Copyright © Huawei Technologies Co., Ltd.

- The OLT is connected to the ONT through a PON board. An MD is deployed for the service channel between the upstream port (0/5/1) of OLT and the user port of ONT (or modem). The MD has a higher level than that of the previous MD.
- The DSLAM connects to the DSL modem to provide various services to users.

Figure 5-24 Typical networking of CFM OAM



Data Plan

Table 5-5 shows the key data plan for deploying the MD for management channel between cascading port on OLT and upstream port on DSLAM.

Table 5-5 Data plan for the Ethernet CFM OAM - MD level 0

Item	Data
OLT	• MD ID: 0. MD name: fttc_md0 • MD level: 0 • MA ID: 0. MA name: fttc_ma0 • MA VLAN: 100 (Management VLAN for the OLT) • MEP ID: 1. MEP port: 0/2/1 • MEP VLAN tag: 8 (Management VLAN for the DSLAM). MEP direction: Down • Remote MEP ID: 2 • CC-interval: 10 minutes

 Copyright © Huawei Technologies Co., Ltd.

Item	Data
DSLAM	• MD ID: 0. MD name: fttc_md0 • MD level: 0 • MA ID: 0. MA name: fttc_ma0 • MA VLAN: 8 (Management VLAN for the DSLAM) • MEP ID: 2. MEP port: 0/8/1 • MEP direction: Down • Remote MEP ID: 1 • CC-interval: 10 minutes

Table 5-6 shows the key data plan for deploying another MD for service channel between upstream port on OLT and UNI port of ONT (or modem).

Table 5-6 Data plan for the Ethernet CFM OAM - MD level 1

Item	Data
OLT	• MD ID: 1; MD name: fttc_md1 • MD level: 1 • MA ID: 1; MA name: fttc_ma1 • MA VLAN: 10 • MEP ID: 1; MEP port: 0/5/1 • MEP VLAN tag1: 10 (VLAN for Internet access services on the upstream port of the upstream board) • MEP direction: Up • Remote MEP ID: 2 • CC-interval: 10 minutes • MAC address: 00e0-fc00-0001
UNI interface of the ONT (or Modem)	• MD ID: 1; MD name: fttc_md1 • MD level: 1 • MA ID: 1; MA name: fttc_ma1 • MA VLAN: 10 • MEP ID: 2; MEP port: 0/1/0 • ONT ID: 0; Ethernet port: 2 • MEP VLAN tag1: 10 • MEP direction: Up • Remote MEP ID: 1 • CC-interval: 10 minutes

Copyright © Huawei Technologies Co., Ltd.

Procedure

- Configure the MD for management channel on the OLT (MD index 0 and MD level 0).

  Pay attention to the following:

  – The MD name and MA name configured on OLT must be the same as that of on DSLAM.
  – Local MEP on OLT corresponds to the remote MEP on the DSLAM, and remote MEP on OLT corresponds to the local MEP on the DSLAM.

  a. Configure the MD.

     Configure MD 0 with name fttc_md0 (character string type), and MD level 0.

     - MDs with the same index or level cannot be created.
     - The name type and the name of an MD must be unique.
     - The total length of the names of an MD and its MAs cannot be longer than 44 characters.
     - The MD name type, the MD name and the MD level must be consistent at both ends.

     ```
     huawei(config)#cfm md 0 name-format string fttc_md0 level 0 mhf-creation no-mhf
     ```

  b. Configure the MA.

     - The system supports up to 4096 MAs. That is, if an MD is configured with 4096 MAs, the other MDs in the system cannot be configured with any MA.
     - An MD of must be available for creating an MA.
     - An existing MA cannot be created again.
     - The total length of the names of an MD and its MAs cannot be longer than 44 characters.
     - The MA name type, the MA name and the sending period of CC packets must be consistent at both ends.

     Create an MA with the index 0/0. The name is fttc_ma0 (character string type). The sending period of CC packets is 10 minutes (the sending period of CC packets is 1 minute by default).

     ```
     huawei(config)#cfm ma 0/0 name-format string fttc_ma0 cc-interval 10m
     ```

     Set the VLAN associated to the MA to 100, it is the management VLAN of the OLT.

     ```
     huawei(config)#cfm ma 0/0 vlan 100
     ```

     Set the ID of MEP contained by the MA to 1 and 2. The number of local MEPs and remote MEPs supported by an MA depends on the type of the service board. Furthermore, all these MEP IDs must be unique. MEP ID 2 needs to be configured on the peer DSLAM.

     ```
     huawei(config)#cfm ma 0/0 meplist 1   //local end MEP
     huawei(config)#cfm ma 0/0 meplist 2 //remote end MEP on the DSLAM
     ```

 Copyright © Huawei Technologies Co., Ltd.

c. Configure the MEP.

- MEP refers to a maintenance association end points. Ethernet CFM OAM is used to test the link connectivity by using the MEPs at the two ends of a maintenance channel.
- By default, the MEP management function is enabled, the priority of sending CFM packets is 7, and the function of sending CC packets is enabled.
- There are two kinds of MEPs: UP MEP and DOWN MEP. An UP MEP indicates that the MEP transmits packets to the bridge trunk direction. A DOWN MEP indicates that the MEP transmits packets to the physical medium direction.
- **vlantag1** or **vlantag2** must be configured, when you add an MEP is added for a port with service streams. **vlantag1** is the outer VLAN of the port carrying the service link for the MEP. **vlantag2** is the inner VLAN of the port carrying the service link for the MEP.
- The MEP priority must be consistent at both ends.

```
huawei(config)#cfm mep 0/0/1 direction down port 0/2/1 vlantag1 8 priority 7
```

d. Enable the remote MEP detection function.

The system can check the remote MEPs of an MA and report alarms for loss of CCM and RDI only when the following functions are enabled: the global CFM function, the global function of checking remote MEPs, and the function of checking the remote MEPs of the MA.

By default, the remote MEP detection function of the MA is enabled, while the global remote MEP detection function is disabled.

i. Enable the remote MEP detection function of the MA.
```
huawei(config)#cfm ma 0/0 remote-mep-detect enable
```
ii. Enable the continuity check function of the MEP.
```
huawei(config)#cfm mep 0/0/1 cc enable
```
iii. Enable the global remote MEP detection function.
```
huawei(config)#cfm remote-mep-detect enable
```

e. Enable the global CFM function.
```
huawei(config)#cfm enable
```

- Configure the MD for management channel on the DSLAM (MD index 0 and MD level 0).

a. Configure the MD.
```
huawei(config)#cfm md 0 name-format string fttc_md0 level 0 mhf-creation no-mhf
```

b. Configure the MA.
```
huawei(config)#cfm ma 0/0 name-format string fttc_ma0 cc-interval 10m
```

Set the VLAN associated to the MA to 8, it is the management VLAN of the DSLAM.

```
huawei(config)#cfm ma 0/0 vlan 8
```

Set the ID of MEP contained by the MA to 2 and 1. The number of local MEPs and remote MEPs supported by an MA depends on the type of the service board. Furthermore, all these MEP IDs must be unique. MEP ID 1 needs to be configured on the peer OLT.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#cfm ma 0/0 meplist 2  //local end MEP on the DSLAM
huawei(config)#cfm ma 0/0 meplist 1 // remote end MEP on the OLT
```

c. Configure the MEP.

```
huawei(config)#cfm mep 0/0/2 direction down port 0/8/1 vlantag1 8 priority 7
```

d. Enable the remote MEP detection function.

i. Enable the remote MEP detection function of the MA.

```
huawei(config)#cfm ma 0/0 remote-mep-detect enable
```

ii. Enable the continuity check function of the MEP.

```
huawei(config)#cfm mep 0/0/2 cc enable
```

iii. Enable the global remote MEP detection function.

```
huawei(config)#cfm remote-mep-detect enable
```

e. Enable the global CFM function.

```
huawei(config)#cfm enable
```

- Configure the MD for service channel on the OLT (MD ID 1 and MD level 1).

a. Configure the MD.

Configure MD 1 with name fttc_md1 (character string type), and MD level 1.

```
huawei(config)#cfm md 1 name-format string fttc_md1 level 1 mhf-creation no-mhf
```

b. Configure the MA.

Create an MA with the index 1/1. The name is fttc_ma1 (character string type). The sending period of CC packets is 10 minutes (the sending period of CC packets is 1 minute by default).

```
huawei(config)#cfm ma 1/1 name-format string fttc_ma1 cc-interval 10m
```

Set the VLAN associated to the MA to 10, it is the service VLAN of the OLT.

```
huawei(config)#cfm ma 1/1 vlan 10
```

Set the ID of MEP contained by the MA to 1 and 2. The number of local MEPs and remote MEPs supported by an MA depends on the type of the service board. Furthermore, all these MEP IDs must be unique. MEP ID 2 needs to be configured on the peer device.

```
huawei(config)#cfm ma 1/1 meplist 1  //local end MEP
huawei(config)#cfm ma 1/1 meplist 2  //remote end MEP on the ONT (or Modem)
```

c. Configure the MEP.

Configure the MEP (direction: up) on a specified port of the upstream board. This port is connected to the user port (SVLAN: 10) of an ONT or modem. If the MEP alarm priority is higher than 1, alarms are reported.

```
huawei(config)#cfm mep 1/1/1 direction up port 0/5/1 vlantag1 10 lowest-alarm-pri 1
```

d. Enable the remote MEP detection function.

The system can check the remote MEPs of an MA and report alarms for loss of CCM and RDI only when the following functions are enabled: the global CFM function, the global function of checking remote MEPs, and the function of checking the remote MEPs of the MA.

By default, the remote MEP detection function of the MA is enabled, while the global remote MEP detection function is disabled.

 Copyright © Huawei Technologies Co., Ltd.

i. Enable the global CFM function.
```
huawei(config)#cfm enable
```
ii. Enable the remote MEP detection function of the MA.
```
huawei(config)#cfm ma 1/1 remote-mep-detect enable
```
iii. Enable the continuity check function of the MEP.
```
huawei(config)#cfm mep 1/1/1 cc enable
```
iv. Enable the global remote MEP detection function.
```
huawei(config)#cfm remote-mep-detect enable
```

e. Configure the alarm indication signal (AIS) function. (Optional.)

i. Set the client level of MA 1/1 to 2.
```
huawei(config)#cfm ma 1/1 client-level 2
```
ii. Enable AIS transmission on MEP 1/1/1.
```
huawei(config)#cfm mep 1/1/1 ais enable
```
iii. Set the interval at which AIS packets of MA 1/1 are sent to 1 minute.
```
huawei(config)#cfm ma 1/1 ais-interval 1m
```

f. Configure the link loss forwarding (LLF) function. (Optional.)

Set CC alarm severity to 3 at which port status association of MEP 1/1/1 is triggered.

```
huawei(config)#cfm mep 1/1/1 trigger if-down 3
```

- Configure the MD for service channel on the ONT (MD ID 1 and MD level 1). ETHOAM CFM configurations of the ONT are completed on the OLT by running commands and are issued to the ONT through the optical network terminal management and control interface (OMCI).

a. Enter the ONT ETHOAM mode.
```
huawei(config)#ont-ethoam
```

b. Configure the MD.

Configure MD 1 with name fttc_md1 (character string type), and MD level 1.

```
huawei(config-ont-ethoam)#cfm md 1 name-format string fttc_md1 level 1
```

c. Configure the MA.

Create an MA with the index 1/1. The name is fttc_ma1 (character string type). The sending period of CC packets is 10 minutes (the sending period of CC packets is 1 minute by default). Set the client level of the MA to 2.

```
huawei(config-ont-ethoam)#cfm ma 1/1 name-format string fttc_ma1 cc-interval 10m
huawei(config-ont-ethoam)#cfm ma 1/1 client-level 2
```

d. Configure the MEP.

Configure the MEP (direction: up) on a specified port of the upstream board. This port is connected to the user port (SVLAN: 10) of an ONT or modem.

```
huawei(config-ont-ethoam)#cfm mep 1/1/2 0/1/0 0 eth 2 direction up vlan 10 remote-mep 1
```

e. Enable the remote MEP detection function.

The system can check the remote MEPs of an MA and report alarms for loss of CCM and RDI only when the following functions are enabled: the global CFM function, the global function of checking remote MEPs, and the function of checking the remote MEPs of the MA.

 Copyright © Huawei Technologies Co., Ltd.

By default, the remote MEP detection function of the MA is enabled, while the global remote MEP detection function is disabled.

i. Enable the global CFM function.
```
huawei(config-ont-ethoam)#cfm 0/1/0 0 enable
```
ii. Enable the remote MEP detection function of the MA.
```
huawei(config-ont-ethoam)#cfm ma 1/1 remote-mep-detect enable
```
iii. Enable the continuity check function of the MEP.
```
huawei(config-ont-ethoam)#cfm mep 1/1/2 cc enable
```
iv. Enable AIS transmission on the MEP.
```
huawei(config-ont-ethoam)#cfm mep 1/1/2 ais enable
```
v. Enable the remote MEP detection function of the ONT.
```
huawei(config-ont-ethoam)#cfm 0/1/0 0 remote-mep-detect enable
```

f. Configure the link LLF function. (Optional)

Set CC alarm severity to 3 at which port status association of MEP 1/1/2 is triggered.

```
huawei(config-ont-ethoam)# cfm mep 1/1/2 trigger if-down 3
```

g. Configure the LB function. (Optional)

Enable the LB test for MEP 1/1/2, set the MAC address of the remote MEP (OLT) to 00e0-fc00-0001, and send 5 LBM packets (data length: 20; priority: 6) at an interval of 1s.

```
huawei(config-ont-ethoam)#cfm loopback mep 1/1/2 dst-mac-address 00e0-fc00-0001 interval 1 count 5 data-length 20 priority 6
```

h. Configure the LT function. (Optional)

Configure the LT function on the link between MEP 1/1/2 and remote destination MEP 1 (OLT).

```
huawei(config-ont-ethoam)#cfm link-trace mep 1/1/2 mepid 1
```

----End

Result

After the configuration is finished,

- OLT or DSLAM is able to learn the MEP ID and MAC address from its remote peer automatically. You can run the **display cfm mep** command to query MEP configuration.
- On the OLT or DSLAM, run the **display cfm mep** command to query the MEP configurations of a specified ONT in the ONT-ETHOAM mode.
- Disconnect OLT and DSLAM, the system will generate CFM OAM alarm automatically, reporting the fault location and cause.
- Run the **cfm loopback** command on the OLT to start a remote loopback test. Under normal circumstances, the number of packets sent and received must be the same.

Configuration File

Configure the MD for management channel on the OLT (MD index 0 and MD level 0).

```
cfm md 0 name-format string fttc_md0 level 0 mhf-creation no-mhf
cfm ma 0/0 name-format string fttc_ma0 cc-interval 10m
```

 Copyright © Huawei Technologies Co., Ltd.

```
cfm ma 0/0 vlan 100
cfm ma 0/0 meplist 1   //local end MEP
cfm ma 0/0 meplist 2 //remote end MEP on the DSLAM
cfm mep 0/0/1 direction down port 0/2/1 vlantag1 8 priority 7
cfm ma 0/0 remote-mep-detect enable
cfm mep 0/0/1 cc enable
cfm remote-mep-detect enable
cfm enable
```

Configure the MD for management channel on the DSLAM (MD index 0 and MD level 0).

```
cfm md 0 name-format string fttc_md0 level 0 mhf-creation no-mhf
cfm ma 0/0 name-format string fttc_ma0 cc-interval 10m
cfm ma 0/0 vlan 8
cfm ma 0/0 meplist 2   //local end MEP
cfm ma 0/0 meplist 1 //remote end MEP on the OLT
cfm mep 0/0/2 direction down port 0/8/1 vlantag1 8 priority 7
cfm ma 0/0 remote-mep-detect enable
cfm mep 0/0/2 cc enable
cfm remote-mep-detect enable
cfm enable
```

Configure the MD for service channel on the OLT (MD ID 1 and MD level 1).

```
cfm md 1 name-format string fttc_md1 level 1 mhf-creation no-mhf
cfm ma 1/1 name-format string fttc_ma1 cc-interval 10m
cfm ma 1/1 vlan 10
cfm ma 1/1 meplist 1   //Indicates the local MEP.
cfm ma 1/1 meplist 2 //Indicates the remote MEP on the DSLAM.
cfm mep 1/1/1 direction up port 0/5/1 vlantag1 10 lowest-alarm-pri 1
cfm enable
cfm ma 1/1 remote-mep-detect enable
cfm mep 1/1/1 cc enable
cfm remote-mep-detect enable
cfm ma 1/1 client-level 2
cfm mep 1/1/1 ais enable
cfm ma 1/1 ais-interval 1m
cfm mep 1/1/1 trigger if-down 3
```

Configure the MD for service channel on the ONT (MD ID 1 and MD level 1).

```
ont-ethoam
cfm md 1 name-format string fttc_md1 level 1
cfm ma 1/1 name-format string fttc_ma1 cc-interval 10m
cfm ma 1/1 client-level 2
cfm mep 1/1/2 0/1/0 0 eth 2 direction up vlan 10 remote-mep 1
cfm 0/1/0 0 enable
cfm ma 1/1 remote-mep-detect enable
cfm mep 1/1/2 cc enable
cfm mep 1/1/2 ais enable
cfm 0/1/0 0 remote-mep-detect enable
cfm mep 1/1/2 trigger if-down 3
cfm loopback mep 1/1/2 dst-mac-address 00e0-fc00-0001 interval 1 count 5 data-length 20 priority 6
cfm link-trace mep 1/1/2 mepid 1
```

5.3.5 EFM (802.3ah)

EFM is defined in the IEEE 802.3ah (802.3ah for short). It is an OAM feature for detecting quality and connectivity of last-mile Ethernet links.

5.3.5.1 EFM Introduction

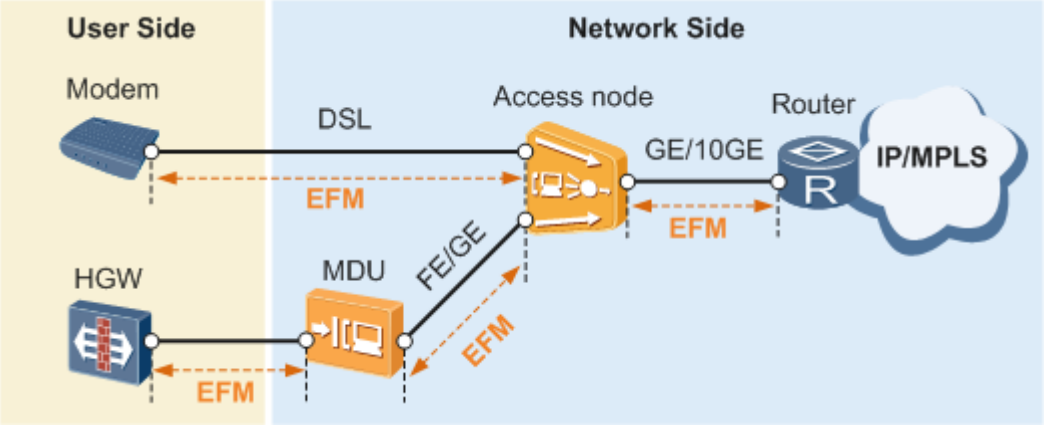
As an OAM mechanism, EFM monitors the link status, quickly locates a faulty link, and identifies the fault type in real time for carriers.

 Copyright © Huawei Technologies Co., Ltd.

Among Ethernet applications, high-end users, such as banks and financial companies, demand high reliability. They expect carriers to monitor both carrier networks and last-mile links that connect users to those carrier networks. However, this is a difficulty for carriers because user devices are out of control of carriers.

Ethernet in the First Mile (EFM) can be used to satisfy these demands. As an OAM mechanism, EFM monitors the link status, quickly locates a faulty link, and identifies the fault type in real time for carriers. EFM mainly applies to the user access network and also to the Ethernet physical link that directly connects two devices. The following example illustrates EFM implementation on the network shown in **Figure 5-25**.

Figure 5-25 Typical EFM network



The router and access node are placed in the CO, MDU is in the corridor or curb, modem and home gateway (HGW) are in users' home. By deploying EFM between the devices shown in **Figure 5-25**, network maintenance engineers can remotely check connectivity and quality of links between these devices.

Table 5-7 lists the main functions of EFM supported by access device.

Table 5-7 Main functions of EFM

Function	Purpose
OAM discovery	Negotiates EFM capabilities of the connected two ends and obtains the EFM status of the remote end.
Remote failure indication	Sends a remote failure indication to the remote end when traffic is interrupted because the device is faulty or is invalid to monitor the link connectivity.
Remote loopback	Loops back the packets sent to the remote end to the local end and tests the link by comparing the received and transmitted packets to obtain the link performance data (including packet loss rate).
Link monitoring	Monitor the link quality by sending a link deterioration event to the network maintenance engineer.

 Copyright © Huawei Technologies Co., Ltd.

5.3.5.2 EFM Basic Concept

Basic concepts of EFM, such as DTE and OAMPDU.

The following concepts are used a lot in the **Principle of EFM**. If you are not familiar with EFM, it is recommended to have a rough understanding of these basic concepts.

DTE

The data terminating entity (DTE) is the basic object on which EFM feature takes effect. A DTE is associated with a port.

- If the port is on the local device, it is a local DTE.
- If the port is on the remote device, it is a remote DTE.

OAMPDUs

EFM works at the data link layer and uses protocol packets called OAM Protocol Data Units (PDUs). EFM DTEs periodically exchange OAMPDUs to report link status, helping network administrators effectively manage networks. **Figure 5-26** shows the OAMPDU format and common types of OAMPDUs. **Table 5-8** lists and describes fields in an OAMPDU.

Figure 5-26 OAMPDU format

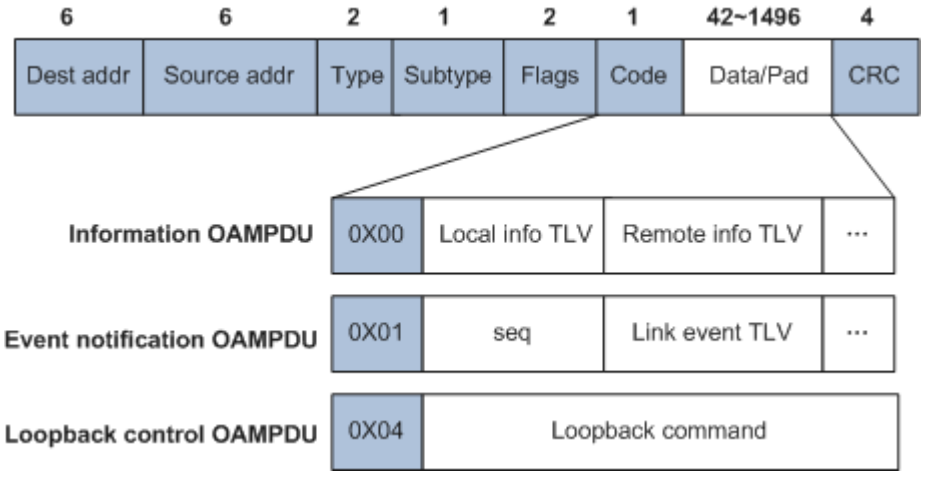


Table 5-8 Fields and descriptions in an OAMPDU

Field	Description
Dest addr	Destination MAC address, which is a slow-protocol multicast address 0x0180-C200-0002. Network bridges cannot forward slow-protocol packets thus EFM OAMPDUs cannot be forwarded over multiple devices.
Source addr	Source address, which is a unicast MAC address of a port on the transmit end. If no port MAC address is specified on the transmit end, the bridge MAC address of the transmit end is used.

 Copyright © Huawei Technologies Co., Ltd.

Field	Description
Type	Slow protocol type, which has a fixed value of 0x8809.
Subtype	Subtype of a slow protocol. The value is 0x03, which means that the slow sub-protocol is EFM.
Flags	Status of a DTE: • Remote Stable • Remote Evaluating • Local Stable • Local Evaluating • Critical Event • Dying Gasp • Link Fault
Code	OAMPDU type: • 0X00: Information OAMPDU • 0X01: Event Notification OAMPDU • 0X04: Loopback Control OAMPDU **Table 5-9** lists common types of OAMPDUs.

Table 5-9 OAMPDU types

OAMPDU Type	Description
Information OAMPDU	• Used to discover a remote DTE, initiate a handshake process, and establish an EFM connection. After the EFM connection is established, both EFM DTEs periodically exchange Information OAMPDUs to monitor link connectivity. • Used to advertise fault information. If the local DTE detects a fault, an event is generated on the local device and an Information OAMPDU is sent to the remote DTE. When the remote DTE receives the Information OAMPDU, an event is generated on the remote device.
Event Notification OAMPDU	Used to monitor links. If a local DTE detects an errored frame event, errored symbol period event, or errored frame second summary event, an event is generated on the local device and an Event Notification OAMPDU is sent to the remote DTE. When the remote DTE receives the Event Notification OAMPDU, an event is generated on the remote device.
Loopback Control OAMPDU	Used to enable or disable the remote loopback function.

Copyright © Huawei Technologies Co., Ltd.

Modes

EFM supports two modes: active and passive. **Table 5-10** lists capabilities for processing OAMPDUs in the two modes.

Table 5-10 Capabilities for processing OAMPDUs in active and passive modes

Capability	Active Mode	Passive Mode
Initiate a connection request by sending an Information OAMPDU during the discovery process	Supported	Not supported
Respond to a connection request during the discovery process	Supported	Supported
Send Information OAMPDUs	Supported	Supported
Send Event Notification OAMPDUs	Supported	Supported
Send Loopback Control OAMPDUs	Supported	Not supported
Respond to Loopback Control OAMPDUs	Supported (The remote DTE must work in active mode.)	Supported

NOTE

- Two DTEs on both ends of an EFM link work in active mode by default.
- An EFM connection can only be initiated by a DTE working in active mode. A DTE working in passive mode waits to receive a connection request from its remote DTE. Two DTEs both working in passive mode cannot establish an EFM connection between them.
- A loopback request can only be initiated by a DTE working in active mode.

5.3.5.3 EFM Principle

EFM functions are defined in section 57 of the IEEE 802.3ah protocol. This topic describes only functions supported by the access device.

NOTE

The following principles are for the access device. Different devices support different functions. For details, see "Feature Specifications and Limitations."

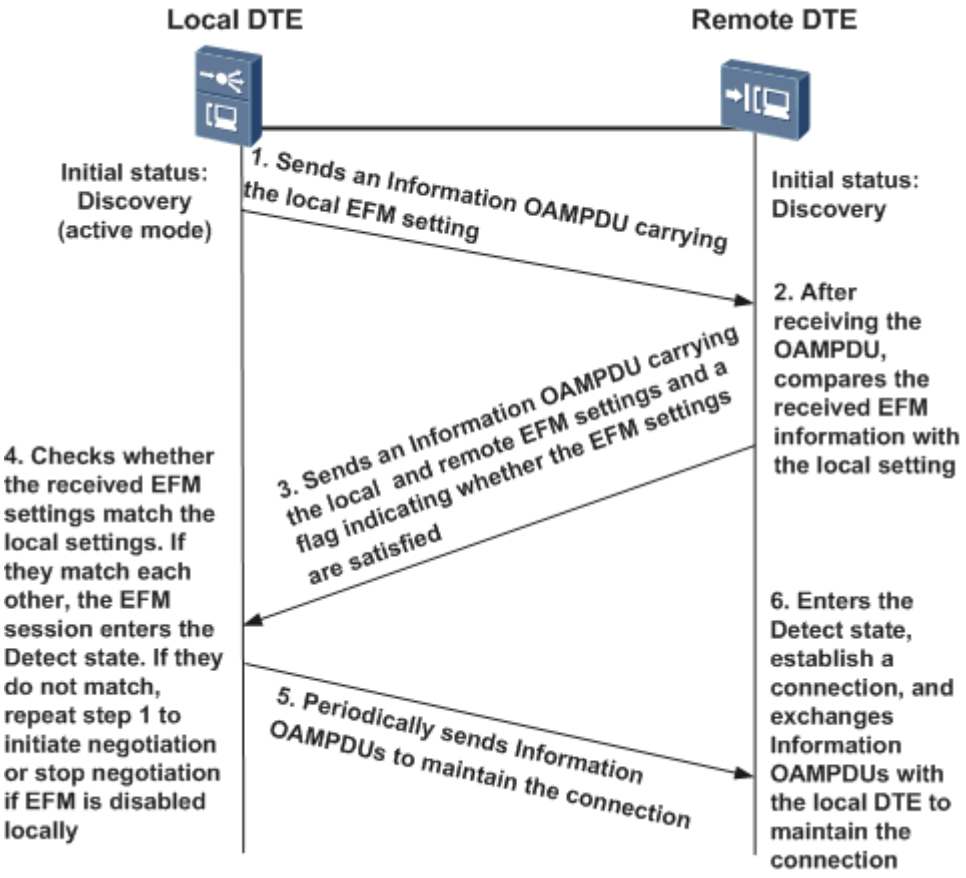
OAM Discovery

After being enabled with EFM, the local DTE initiates the Discovery negotiation. During the Discovery phase, a local DTE discovers and establishes a stable EFM connection with a remote DTE. If both the local and remote DTEs are satisfactory

 Copyright © Huawei Technologies Co., Ltd.

for EFM functions, the EFM connection is established. **Figure 5-27** shows this process.

Figure 5-27 Schematic diagram for OAM discovery



After the EFM connection is established, DTEs at both ends of an EFM connection periodically exchange Information OAMPDUs to monitor link connectivity. The interval at which Information OAMPDUs are sent is also known as an interval between handshakes. If a DTE does not receive Information OAMPDUs from the remote DTE within the connection timeout period, the DTE considers the connection interrupted and generates an event. Establishing an EFM connection is a way to monitor physical link connectivity automatically.

Remote Failure Indication

After the OAM discovery phase finishes, two DTEs at both ends of an EFM connection exchange Information OAMPDUs to monitor link connectivity. If traffic is interrupted due to a remote device failure, the remote DTE sends an Information OAMPDU carrying an event listed in **Table 5-11** to the local DTE. After receiving the notification, the local DTE generates the Ethernet OAM link event. An administrator can determine link status based on the event and take measures to rectify the fault.

 Copyright © Huawei Technologies Co., Ltd.

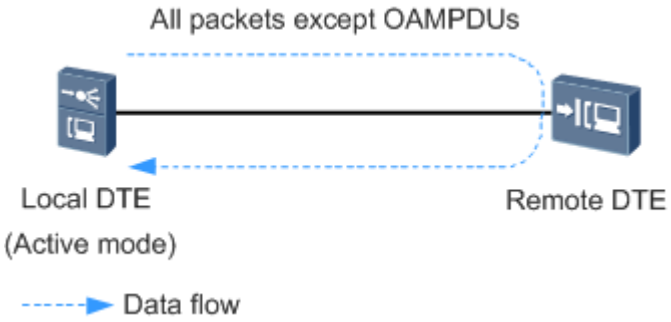
Table 5-11 Critical link events

Type	Description
Link Fault	A loss of signal (LoS) error occurs because the interval at which OAMPDUs are sent elapses or a physical link fails.
Dying Gasp	An unexpected status changes or event occurs because a remote device is powered off or a board is reset.
Critical Event	An unidentified critical event occurs.

Remote Loopback

Figure 5-28 demonstrates the principles of remote loopback. When a local DTE sends non-OAMPDUs to a remote DTE, the remote DTE loops the non-OAMPDUs back to the local DTE, not to the destination addresses of the non-OAMPDUs. This is remote loopback. It helps to obtain the link performance data (including packet loss rate) by comparing the received and transmitted packets.

Figure 5-28 Principles of EFM remote loopback

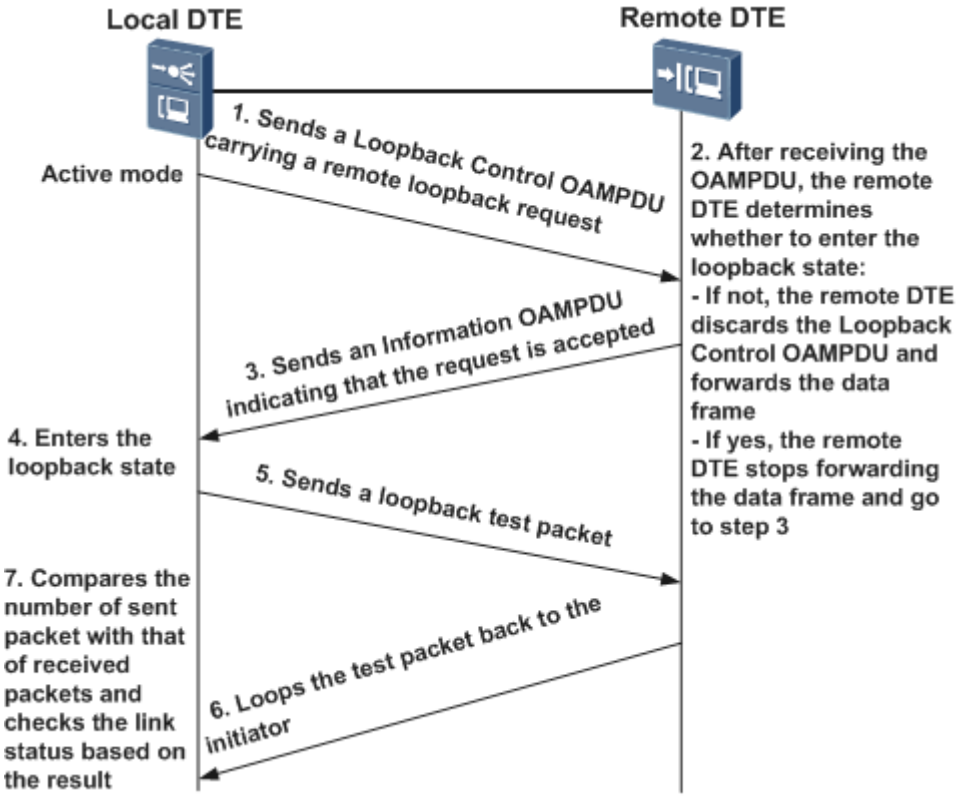


To implement remote loopback, an EFM connection must be established first. After remote loopback is enabled, all data frames forwarding except OAMPDUs are stopped and services will be interrupted. This function can be used to detect the link quality before a new network is deployed or after a link fault is rectified to prevent services from being affected.

In remote loopback mode, the local DTE sends testing packets to the remote DTE. The local DTE then computes communication quality parameters, such as the packet loss ratio, of the current link according to the numbers of sent packets and received packets. If the local DTE attempts to stop remote loopback, it sends a message to instruct the remote DTE to disable remote loopback. After receiving the message, the remote DTE disables remote loopback. **Figure 5-29** shows the remote loopback process.

Copyright © Huawei Technologies Co., Ltd.

Figure 5-29 Remote loopback process



Link Monitoring

Monitoring Ethernet links is difficult if network performance deteriorates while traffic is being transmitted over physical links. To resolve this problem, the EFM link monitoring function can be used. It can detect data link layer faults in various environments. DTEs that are enabled with link monitoring exchange Event Notification OAMPDUs to monitor links.

After monitoring common link events listed in **Table 5-12**, the DTE at an end sends OAMPDUs carrying corresponding events to the other end and generates Ethernet OAM link events at local. When the other end receives the OAMPDUs, it generates an event. This facilitates real-time network monitoring and effective network management for maintenance engineers.

Table 5-12 Common link events

Type	Description	Purpose
Errored Symbol Period Event	The number of errored symbols during a period reaches the threshold.	To detect code errors during data transmission at the physical layer.
Errored Frame Event	The number of frames during the unit time reaches the preset threshold.	To detect frame errors that occur during data transmission at the MAC sublayer.

 Copyright © Huawei Technologies Co., Ltd.

Type	Description	Purpose
Errored Frame Period Event	The number of errored frames in the latest N frames reaches the threshold.	
Errored Frame Seconds Summary Event	The number of errored frames in the latest M seconds reaches the threshold. **NOTE** If an errored frame is generated at a second, this second is called errored frame second.	To detect errored frame seconds that occur during data transmission at the MAC sublayer.

5.3.5.4 EFM Configuration

5.3.5.4.1 Configuring an EFM Remote Loopback

In an Ethernet in the first mile (EFM) remote loopback, all non-OAM protocol data units (PDUs) sent from the local data terminating entity (DTE) to the remote DTE are looped back to the local DTE. The link performance data (such as packet loss rate) can be obtained by comparing the sent and received non-OAM PDUs on the local DTE. This helps to evaluate Ethernet link quality and locate faults.

Context

Before the EFM remote loopback starts, the EFM feature must be enabled on the local DTE and the remote DTE. After the EFM feature is enabled, the EFM mode and the response mode of EFM remote loopback cannot be modified on the port. If these two modes need to be modified, disable the EFM feature.

An EFM remote loopback results in service interruption. Therefore, stop the loopback right after the link performance data is collected.

The following describes how to configure an EFM remote loopback by assuming that the local DTE initiates an EFM remote loopback and the remote DTE responds to the EFM remote loopback.

Procedure

Step 1 Configure the remote DTE.

1. Run the **efm loopback frameid/slotid/portid process** command to set the response mode of EFM remote loopback on a port. By default, the response mode of EFM remote loopback on a network-side port is set to process, while the response mode of EFM remote loopback on a user-side port is set to ignore.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

Only a few boards can respond to an EFM remote loopback. For these boards, run the above command to set the response mode to process or ignore, that is, to respond to or ignore the EFM remote loopback initiated by the peer DTE. For boards that cannot respond to an EFM remote loopback, the response mode is defaulted to ignore and cannot be modified.

2. Run the **efm oam frameid/slotid/portid enable** command to enable the EFM feature for the port. By default, the EFM feature is disabled.

Step 2 Configure the local DTE.

1. Run the **efm oam mode frameid/slotid/portid active** command to set the EFM mode of a port to the default value active.
2. Run the **efm oam frameid/slotid/portid enable** command to enable the EFM feature for the port. By default, the EFM feature is disabled.
3. Run the **efm loopback frameid/slotid/portid start** command to start the EFM remote loopback. By default, the EFM remote loopback stops.

NOTE

If you want the EFM remote loopback to automatically stop after a period of time, configure the **timeout** parameter to set the loopback timeout period when running the **efm loopback frameid/slotid/portid start** command.

----End

Result

All packets except OAM PDUs sent from the local DTE to the remote DTE are looped back to the local DTE.

Example

Assume that Ethernet cascade port 0/2/0 on the OLT is connected to upstream port 0/0/1 on the MDU, Ethernet cascade port on the OLT initiates an EFM remote loopback, and upstream port on the MDU responds to the EFM remote loopback. To configure an EFM remote loopback on the two ports, do as follows:

1. Configure MDU.
   MDU(config)#**efm oam 0/0/1 enable**
2. Configure OLT.
   OLT(config)#**efm oam 0/2/0 enable**
   OLT(config)#**efm loopback 0/2/0 start**

Follow-up Procedure

After the link performance data is collected, run the **efm loopback frameid/slotid/portid stop** command on the local DTE to stop the EFM remote loopback.

NOTE

If the timeout period has been set, the EFM remote loopback will automatically stop after the timeout period expires.

 Copyright © Huawei Technologies Co., Ltd.

5.3.5.4.2 Configuring EFM Ethernet Link Monitoring

The Ethernet in the First Mile (EFM) is an operation, administration and maintenance (OAM) feature which provides a real-time link monitoring mechanism. Using this mechanism, events are reported in the case of traffic interruption or link quality degradation so that the network maintenance engineers can monitor Ethernet link quality in real time.

Context

EFM events can be classified into two types: local EFM events and remote EFM events. Local EFM events are detected by the local DTE; remote EFM events are detected by the remote DTE and sent to the local DTE.

- All remote EFM events will be reported as long as the EFM feature is enabled. There is no command that can be used to configure the event reporting function for the remote EFM events.
- For some local EFM events, you need to enable the event reporting function so that the local EFM events can be reported. As listed in the **Table 5-13**, the local DTE reports the related EFM events and notifies the remote DTE only when the EFM feature and the event reporting function are enabled. If the EFM feature is enabled but the event reporting function is disabled, you can run the **display efm oam event current frameid/slotid/portid local** command to query the EFM events detected by the local DTE.

  Table 5-13 Conditions for detecting and reporting local EFM events

EFM Feature	Event Reporting Function	Detect Local EFM Events or Not	Report Local EFM Events or Not
Enabled	Enabled	Yes	Yes
	Disabled	Yes	No
Disabled	Enabled	No	No
	Disabled	No	No

The local DTE detects only three types of EFM events: link fault event, errored frame event, and errored frame seconds summary event.

- For the errored frame event and errored frame seconds summary event, the event reporting function is disabled by default. You need to run some commands to enable this function. For the specific commands, see the following configuration steps.
- For the link fault event, the event reporting function is enabled by default and cannot be modified.

Procedure

Step 1 Run the **efm oam frameid/slotid/portid enable** command to enable the EFM feature for the port. By default, the EFM feature is disabled.

 Copyright © Huawei Technologies Co., Ltd.

Step 2 (Optional) Configure the event reporting function and related parameters on the local DTE.

When the local DTE needs to report the errored frame event or the errored frame seconds summary event and notify the remote DTE, perform this step by running the **efm error-frame** command or **efm error-frame-second** command respectively.

Step 3 Query the configuration results. The following table lists the query commands related to EFM Ethernet link monitoring.

Item	Command
Querying the status of the EFM feature on the local DTE	**display efm oam status frameid/slotid/portid local** **NOTE** In the command output, the **Admin Status** parameter indicates whether the EFM feature is enabled on the local DTE.
Querying the status of the remote DTE	**display efm oam status frameid/slotid/portid remote** **NOTE** In the command output, the **Event Support** parameter indicates whether the remote DTE supports sending a notification to the local DTE when the remote DTE detects an EFM event.
Querying the event reporting function and related parameters of different EFM events	**display efm oam event config**
Querying the latest EFM event	**display efm oam event current**
Querying the number of EFM events	**display statistics performance**

----End

Result

In the case of traffic interruption or link quality degradation, the access device (local DTE) reports events based on the configurations.

Example

Assume that Ethernet cascade port 0/2/0 on the OLT is connected to upstream port 0/0/1 on the MDU, both OLT and MDU can receive the EFM events sent by each other and can detect EFM events, and the detection interval and other thresholds are defaulted. To monitor the quality of this Ethernet link section, do as follows:

- OLT configurations:
  ```
  OLT(config)#efm oam 0/2/0 enable
  OLT(config)#efm error-frame 0/2/0 notification enable
  OLT(config)#efm error-frame-second 0/2/0 notification enable
  ```

 Copyright © Huawei Technologies Co., Ltd.

- MDU configurations:

```
MDU(config)#efm oam 0/0/1 enable
MDU(config)#efm error-frame 0/0/1 notification enable
MDU(config)#efm error-frame-second 0/0/1 notification enable
MDU(config)#efm error-frame-period 0/0/1 notification enable
```

NOTE

Of the OLT and MDU, only the MDU can detect the errored frame period event, and only MDUs of some models can detect this event. For details, see the *Feature Specifications* in related MDU product documents.

5.3.5.5 EFM Maintenance and Diagnosis

After EFM is enabled, the access device uses **0x2f00000a Ethernet OAM Link Events** to implement **remote failure indication** and **link monitoring**. According to event parameters, the event type (such as link fault and errored frame) and location (local or remote) can be determined.

- The access device can receive all types of EFM events sent by the remote device and then report **0x2f00000a Ethernet OAM Link Events** at local.
- The access device can only detect some types of EFM events, report **0x2f00000a Ethernet OAM Link Events** at local, and uses the OAMPDU to notify the remote end. Types of EFM events can be detected at local are related to the hardware. For the hardware supported, see "Specifications."

After the EFM connection is established, the remote DTE status can be queried by the **display efm oam status** command.

The statistics of EFM OAM packets that are received and sent by local TDE can be queried by the **display efm oam statistics** command.

5.3.6 PM (Y.1731)

Performance monitoring (PM) is defined in ITU-T Y.1731. PM, as an OAM feature, is used to test the performance of Ethernet links and evaluate the network quality.

5.3.6.1 PM Introduction

This topic describes the benefits and functions of performance monitoring (PM) defined in ITU-T Y.1731.

Benefits

As the multi-play service (for example, IPTV) prevails, more and more difficulties in Ethernet OAM emerge. The following uses the triple-play service as an example to explain the difficulties facing carriers.

- IPTV: Video streams are sensitive to packet loss and delay. The loss of a packet affects dozens of pictures, causing frame freezing and impairing user experience.
- VoIP: Audio streams are sensitive to packet loss and delay. Packet loss or overlong delay results in intermittent voices; delay variations result in vague voices, which are intolerable to users.

 Copyright © Huawei Technologies Co., Ltd.

- High-speed Internet (HSI): Packet loss or delay occurring in the data streams may cause user authentication failures or user connection failures, which increases user complaints.

As a result, minimizing user complaints (caused by network quality degradation) and improving fault locating efficiency have become carriers' top concerns.

Besides, in the case of national broadband, carriers charge retail service providers (RSPs) based on the level of network quality when selling bandwidth to the RSPs. Some carriers even charge end users based on the level of network quality. Therefore, carriers need to monitor the network quality and provide the service level agreement (SLA) as the accounting basis.

ITU-T Y.1731 defines a series of methods for monitoring Ethernet link performance. These methods help carriers collect link performance data in real time and fast diagnose the network performance faults. ITU-T Y.1731 also provides carriers with technical means for outputting the SLA reports.

Functions

The following table lists the three PM functions supported by access devices.

Function	Description
Single-ended ETH-LM	Single-ended Ethernet loss measurement (ETH-LM) is used to collect the packet loss rate of an Ethernet link based on a count of transmitted and received data frames between a pair of MEPs.
Two-way ETH-DM	Two-way Ethernet delay measurement (ETH-DM) is used to collect the delay, delay variation of an Ethernet link.
Single-ended ETH-SLM	Single-ended Ethernet synthetic loss measurement (ETH-SLM) is a mechanism to measure frame loss using synthetic frames, rather than data traffic. A number of synthetic frames are sent and received, and the number of those that are lost is hence calculated. This can be treated as a statistical sample, and used to approximate the frame loss ratio of data traffic .

Both single-ended ETH-LM, two-way ETH-DM and single-ended ETH-SLM support the following modes:

- On-demand mode: In this mode, performance data is collected in real time. On-demand single-ended ETH-LM and two-way ETH-DM are used for diagnosing network quality faults. They can be started through the CLI or NMS, and related measurement results are displayed on the CLI or NMS.
- Periodic mode: In this mode, performance data is collected and reported periodically. Periodic single-ended ETH-LM and two-way ETH-DM are used by carriers to output periodic SLA reports. They can be started only through the NMS.

 Copyright © Huawei Technologies Co., Ltd.

5.3.6.2 PM Networking Application

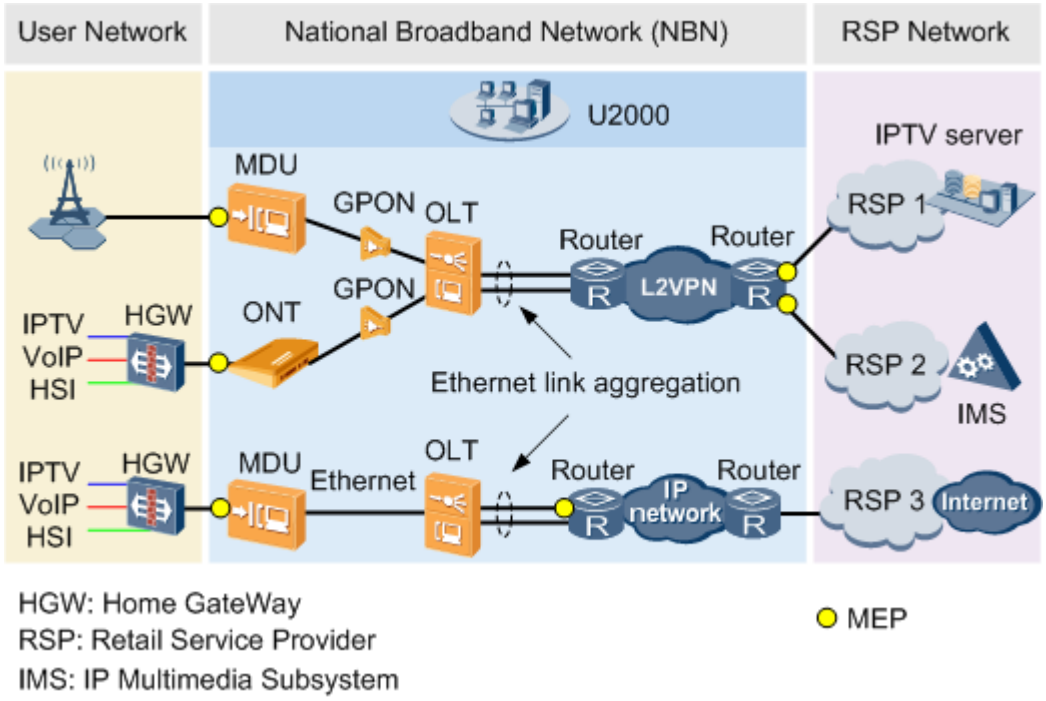
Performance monitoring (PM) is mainly used on the national broadband (NBB) network to monitor and evaluate carriers' service quality. In addition, some carriers deploy maintenance association end points (MEPs) on their networks. These MEPs help carriers test link performance section by section along a network link, achieving fast fault demarcation and fault locating.

The following describes the typical MEP deployment on common networks.

NBB Network

Figure 5-30 shows the typical MEP deployment on the NBB network to implement PM according to ITU-T Y.1731. The user ports on the MDUs and ONTs mark the boundary between the NBB network and the user network; the network-side ports on routers mark the boundary between the NBB network and the retail service provider (RSP) network. MEPs deployed on the boundaries of the NBB network collect the performance data about the Ethernet links between 2 MEPs. The collected performance data is used for network quality evaluation. ITU-T Y.1731 is a Layer 2 network protocol, and the Ethernet protocol packets defined by ITU-T Y.1731 cannot traverse an IP network. If MEPs need to be deployed on the network-side ports on the routers at the boundary of the NBB network, a Layer 2 VPN tunnel needs to be established between the routers.

Figure 5-30 Typical MEP deployment on the NBB network



NOTE

Currently, no ONT supports Ethernet synthetic loss measurement (ETH-SLM) defined in ITU-T Y.1731.

Copyright © Huawei Technologies Co., Ltd.

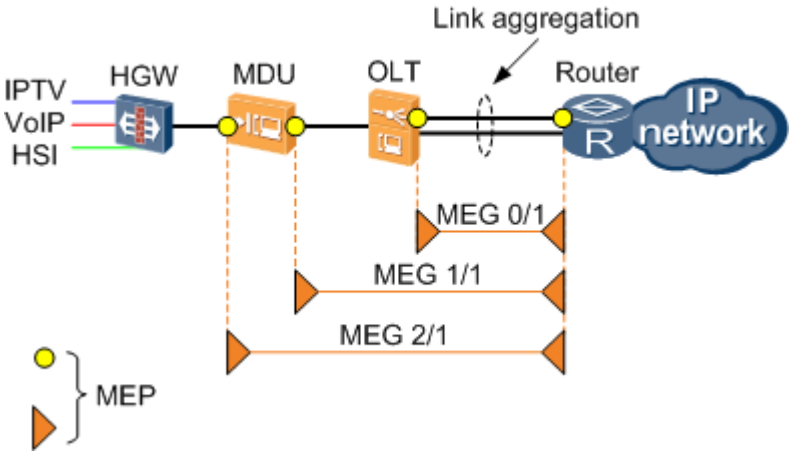
FTTB/FTTC Network

Figure 5-31 shows the MEPs deployed on the user port and upstream port on the MDU, upstream link aggregation group of the OLT, and downstream link aggregation group of the router on a fiber to the building (FTTB) or a fiber to the curb (FTTC) network. These MEPs help test the link performance in each maintenance entity group (MEG).

For example, if a user reports that the voice is intermittent during the call, maintenance engineers perform single-ended Ethernet loss measurement (ETH-LM) section by section along the network link to diagnose the specific network section in which the user's voice service stream encounters packet loss. As shown in Figure 5-31, maintenance engineers check whether packets are lost in MEG 2/1. If no packet is lost, the link between the user port on the MDU and the router runs normally. If packets are lost, maintenance engineers continue to check whether packets are lost in MEG 1/1.

- If packets are lost in MEG 1/1, maintenance engineers check whether packets are lost in MEG 0/1.
- If no packet is lost in MEG 1/1 but packets are lost in MEG 2/1, it indicates that packets are lost on the MDU.

Figure 5-31 Typical MEP deployment on an FTTB/FTTC network

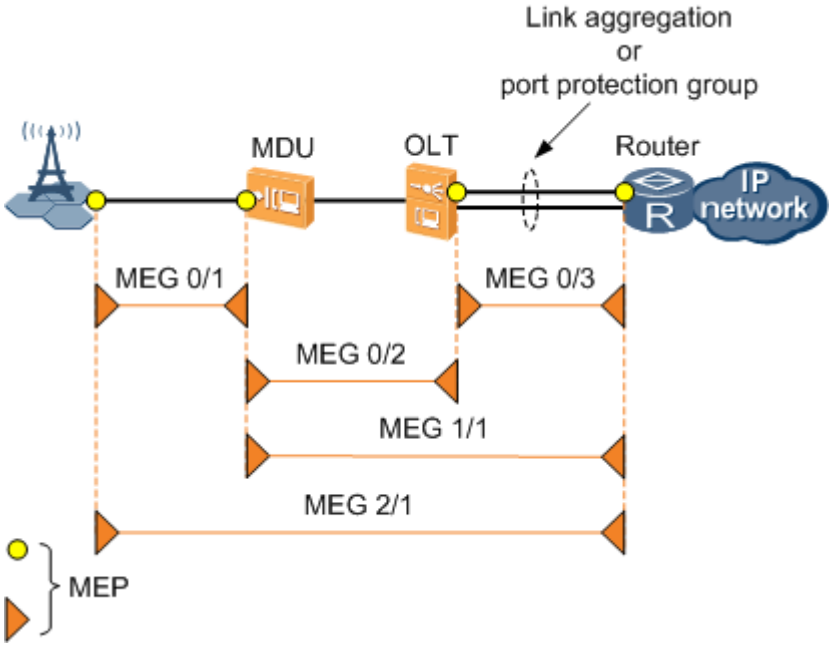


FTTM Network

Fiber to the mobile base station (FTTM) networks have a high requirement on link reliability. Therefore, port protection groups or link aggregation groups (active/standby mode) are created on the links between the OLT and router on an FTTM network. Figure 5-32 shows the MEPs deployed on the upstream port on the mobile base station, user port on the MDU, upstream port protection group or link aggregation group of the OLT, and user-side port protection group or downstream link aggregation group of the router. These MEPs help test the link performance in each MEG.

Copyright © Huawei Technologies Co., Ltd.

Figure 5-32 Typical MEP deployment on an FTTM network

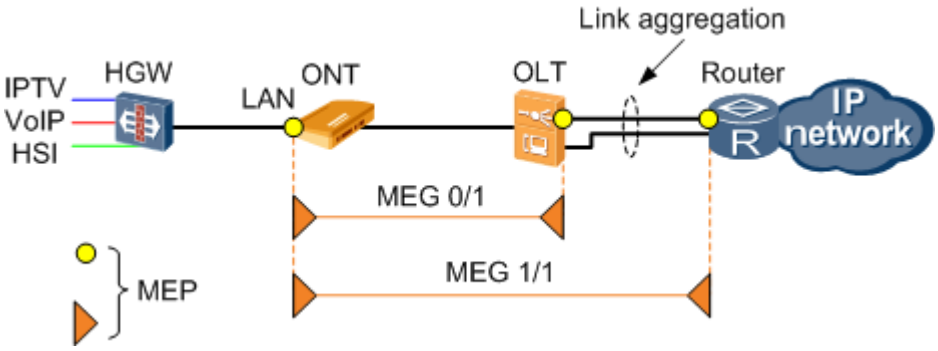


FTTH Network

The MEP is deployed on the LAN port or the WAN port on the ONT, upstream link aggregation group of the OLT, and downstream link aggregation group of the router. These MEPs help test the link performance in each MEG.

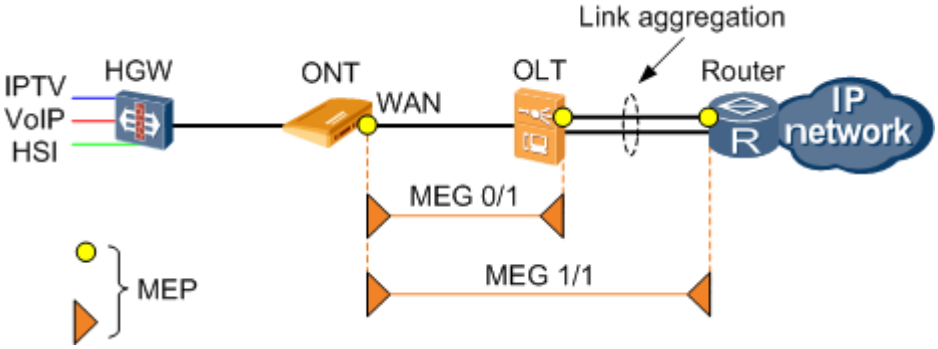
- **Figure 5-33** shows the up MEP deployed on the LAN port of a bridging ONT.
- **Figure 5-34** shows the down MEP deployed on the WAN port of a routing ONT.

Figure 5-33 Typical MEP deployment on an FTTH network with bridging ONT



Copyright © Huawei Technologies Co., Ltd.

Figure 5-34 Typical MEP deployment on an FTTH network with routing ONT



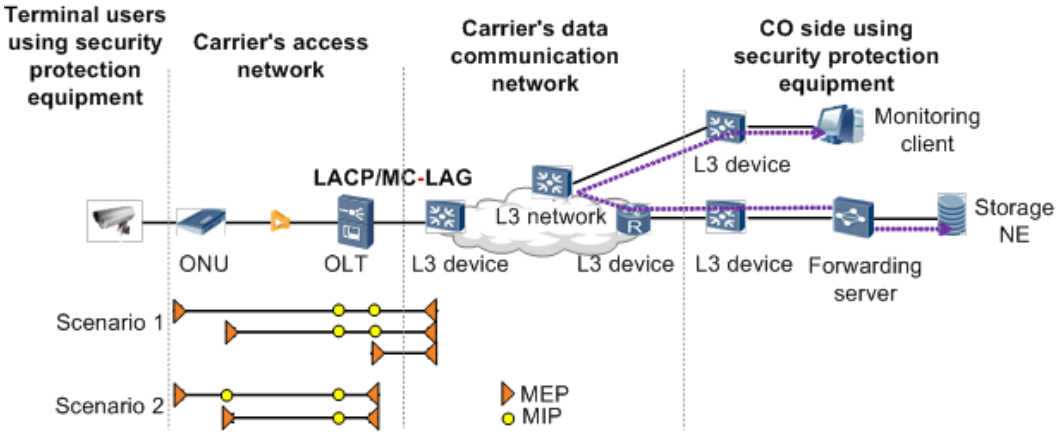
In FTTH scenario, the performance monitoring function of ONT is configured through the CLI of OLT. Then OLT issues the configuration to ONT through the optical network terminal management and control interface (OMCI) channel. ONT can only function as the responding end.

FTTO Video Monitoring Network

On-demand video monitoring is preferred for fault demarcation, as shown in **Figure 5-35**.

- Scenario 1. MEGs are deployed on the Layer 3 UNI and ONU UNI to check continuity and check packet loss, delay, and jitter in one direction. In this scenario, the protection network between the upstream port on the OLT and Layer 3 device needs to be considered, including single-homing Link Aggregation Control Protocol (LACP) protection and dual-homing multi-chassis link aggregation group (MC-LAG) protection.
- Scenario 2. MEGs are deployed on the OLT NNI and ONU UNI to check continuity and check packet loss, delay, and jitter in one direction.

Figure 5-35 Typical MEP deployment on an FTTO video surveillance network



5.3.6.3 PM Basic Concepts

This topic introduces the basic concepts of performance monitoring (PM) defined in ITU-T Y.1731.

Some concepts are detailed in ITU-T Y.1731 and IEEE 802.1ag. When access devices implement PM, some concepts, such as the maintenance association end

 Copyright © Huawei Technologies Co., Ltd.

point (MEP) and maintenance association (MA) are used. For details, see **5.3.4.3 CFM Basic Concepts** and **5.3.3 Differences in Implementing Y.1731 and 802.1ag on Access Device**.

Besides, Huawei introduces some concepts (such as the PM instance and PM task) that are not defined in the protocols. The following table lists the items that need to be configured during PM deployment. The relationships between the configured items are shown in **Figure 5-36**.

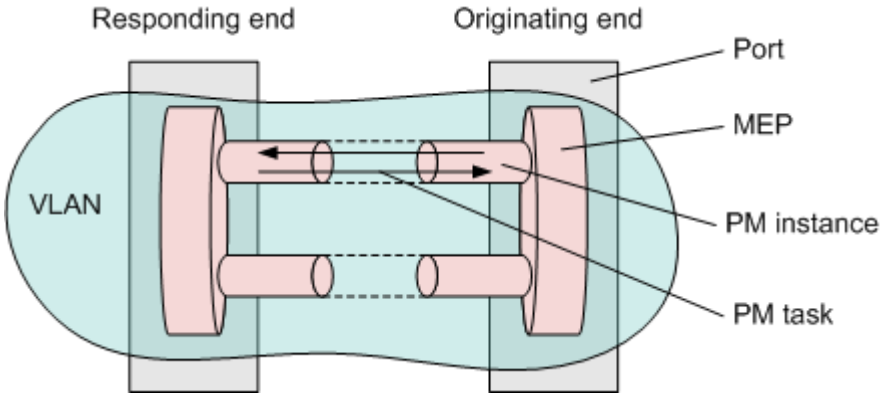
Item	Function
MEP	Used to demarcate the Ethernet link to be monitored. MEPs can be divided into originating and responding ends based on the type of Ethernet protocol packets sent by them. • Originating end: sends the Ethernet protocol request packets and receives the Ethernet protocol response packets sent by the responding end. By comparing the request and response packets, the system calculates the packet loss rate, number of lost packets, and the delay occurring on the link between the 2 MEPs. • Responding end: replies to the originating end with Ethernet protocol response packets. The responding end does not calculate or save the performance test results.

 Copyright © Huawei Technologies Co., Ltd.

Item	Function
PM instance	Used to select the PM object on the Ethernet link between 2 MEPs. PM instances are identified by the test ID. The following parameters are provided to differentiate between different PM instances. • **mep**, **remote-mep**, and **remote-mep-mac**: specify the two ends of the Ethernet link to be monitored. Two MEPs exchange Ethernet protocol packets with each other to collect performance data. • **receive-8021p** and **send-8021p**: specify the priority of Ethernet protocol packets sent and received by the MEP. The priority of the Ethernet protocol packets must be the same as that of the to-be-monitored service flow. In **Multi-priority Convergence** scenario, to calculate all the packets received by or sent from the MEP according to Huawei's proprietary protocol, set these parameters to **unaware**. Parameter value **unaware** is valid to Ethernet loss measurement (ETH-LM), Ethernet delay measurement (ETH-DM), and Ethernet synthetic loss measurement (ETH-SLM). • **backward-mac**, **onward-mac**, **update-backward-mac**, and **update-onward-mac**: defined by Huawei's proprietary protocol. These 4 parameters are used to differentiate between service flows when an ETH-LM or ETH-DM test is performed in the **point-to-multipoint connection (N:1 VLAN)** scenario. These 4 parameters do not need to be configured when an Ethernet synthetic loss measurement ETH-SLM is performed or when an ETH-LM or ETH-DM test is performed in the **point-to-point connection (1:1 VLAN)** scenario.
PM task	Used to select one of the PM instances configured in the system for an LM or delay measurement (DM) test. PM instances are configured manually and do not trigger the Ethernet protocol packet exchange between the MEPs at both ends of the link. The 2 MEPs normally exchange Ethernet protocol packets only when the originating and responding ends use the same PM instance for performing the same PM task. • **pm lm send** and **pm dm send**: used to configure the interval for sending Ethernet protocol packets (used for LM and DM tests) and configure the number of Ethernet protocol packet sending attempts on the originating end. After these 2 parameters are configured, the originating end starts to send Ethernet protocol request packets. • **pm lm receive**, **pm dm receive**, and **pm slm receive**: used to configure whether the responding end responds to the requests initiated by the originating end.

 Copyright © Huawei Technologies Co., Ltd.

Figure 5-36 Items that need to be configured during PM deployment



5.3.6.4 PM Principles

Performance monitoring (PM) defined in ITU-T Y.1731 cannot be used in some application scenarios. Therefore, Huawei broadens the PM application scenarios based on Huawei's proprietary protocol. The following introduces how to implement Ethernet link PM on an FTTB/FTTC network with typical VLAN and QoS planning.

Access devices implement single-ended Ethernet loss measurement (ETH-LM), two-way Ethernet delay measurement (ETH-DM) and single-ended Ethernet synthetic loss measurement (ETH-SLM) according to ITU-T Y.1731. For basic implementation principles, see ITU-T Y.1731.

5.3.6.4.1 Performance Monitoring with Typical VLAN Planning

Single-ended Ethernet loss measurement (ETH-LM) and two-way Ethernet delay measurement (ETH-DM) defined in ITU-T Y.1731 are based on the point-to-point (P2P) connection. That is, a maintenance association end point (MEP) receives packets only from a remote MEP. However, in actual networking applications, the VLAN planning is very flexible, which means that a MEP receives packets from multiple remote MEPs. This networking mode is a type of point-to-multipoint (P2MP) connection mode. The following introduces how to implement single-ended Ethernet loss measurement (ETH-LM) and two-way Ethernet delay measurement (ETH-DM) in P2P and P2MP connection scenarios with typical VLAN planning.

 Copyright © Huawei Technologies Co., Ltd.

P2P Connection (1:1 VLAN Scenario)

Figure 5-37 P2P network for enterprise users

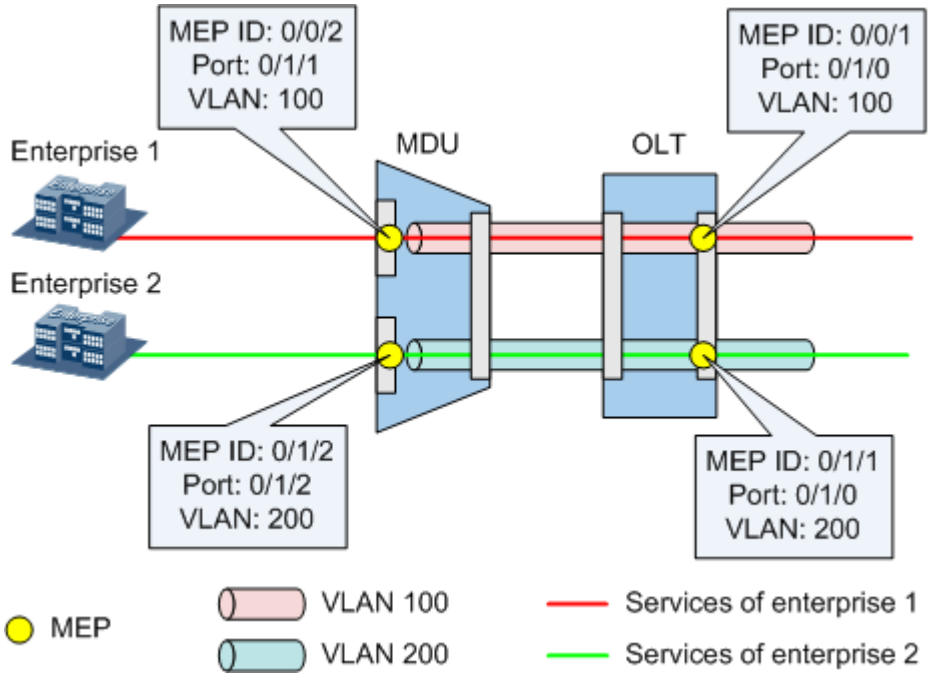


Figure 5-37 shows the typical 1:1 VLAN planning. The MDU translates the VLAN of the enterprise user's service flow to the service VLAN (S-VLAN). Then the OLT transparently transmits the VLAN. The following conditions must be met during the S-VLAN planning.

- Different enterprise users connected to the same OLT must be configured with different S-VLANs.
- Packets sent by any 2 MDUs connected to the same OLT cannot carry the same S-VLAN tag.

Assume that enterprise users' services in Figure 5-37 are transmitted upstream by the upstream interface board on the OLT. If maintenance engineers need to check the performance of the link between the MDU user port and the OLT upstream port, deploy the MEPs on the MDU user port and the OLT upstream port as shown in Figure 5-37. All the packets received on MEP 0/0/1 are sent from MEP 0/0/2. Such a connection is a P2P connection. For a P2P connection, when a PM instance is added, the MEPs at both ends of the link and the priorities of to-be-test services must be specified.

 Copyright © Huawei Technologies Co., Ltd.

P2MP Connection (N:1 VLAN Scenario)

Figure 5-38 P2MP network for home users

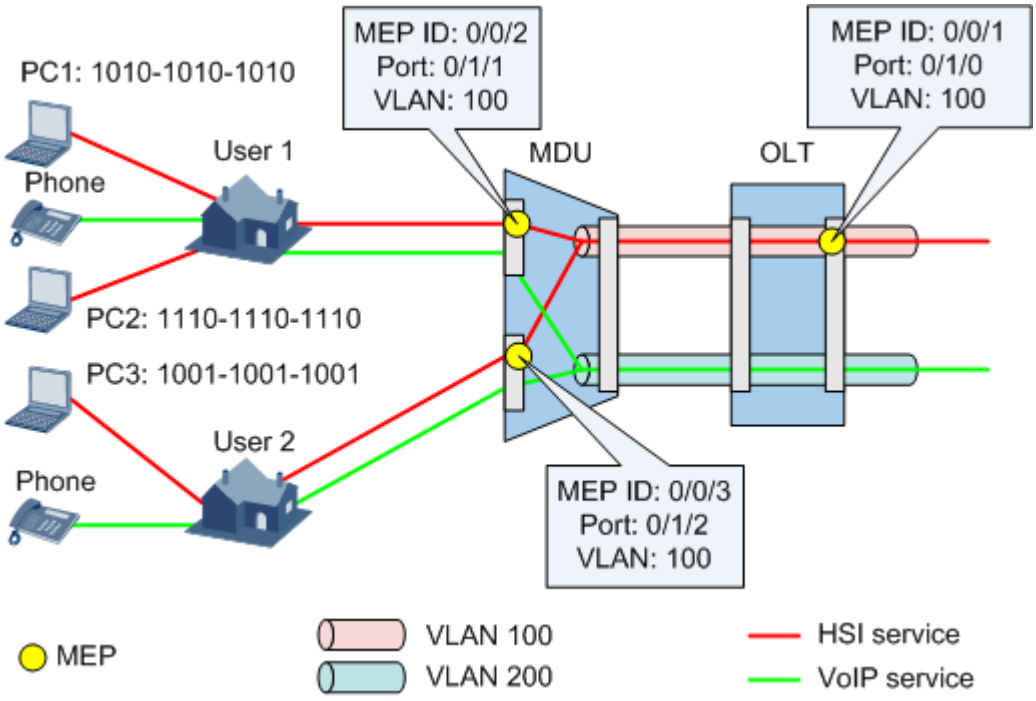


Figure 5-38 shows the typical N:1 VLAN planning. The MDU translates the VLAN of the home user's service flow to the S-VLAN. Then the OLT transparently transmits the VLAN. The following conditions must be met during the VLAN planning.

- Customer VLANs (C-VLANs) are used to differentiate between home users' service flows.
- The MDU translates C-VLANs to S-VLANs. Packets of the same type are transmitted upstream through the same S-VLAN.

Assume that home users' services in Figure 5-38 are transmitted upstream through the upstream interface board on the OLT. If maintenance engineers need to measure the packet loss rate of user 1's HSI services on an FTTx network, deploy the MEPs on the MDU user port and the OLT upstream port as shown in Figure 5-38. All data services on the MDU are transmitted upstream through the same S-VLAN. All the packets received on MEP 0/0/1 are sent from MEP 0/0/2 and MEP 0/0/3. Such a connection is a P2MP connection. In this case, MEP 0/0/1 cannot differentiate between packets from MEP 0/0/2 and those from MEP 0/0/3. Therefore, the packet loss rate cannot be accurately measured on the link between MEP 0/0/1 and MEP 0/0/2. To solve this issue, Huawei defines 2 concepts, onward MAC and backward MAC, in its proprietary protocol. Onward MAC and backward MAC are used to identify P2P service flows on a P2MP connection network.

Essentially, onward MAC and backward MAC are MAC addresses of user terminals. Each MAC address is globally unique. Therefore, the MEP where service flows converge can identify the source of the packets based on the MAC address.

- Backward MAC: used to configure the PM instance at the access end (for example, the MDU user port in Figure 5-38) of the service flows.

 Copyright © Huawei Technologies Co., Ltd.

- Onward MAC: used to configure the PM instance at the convergence end (for example, the OLT upstream port in **Figure 5-38**) of the service flows.

Onward MAC and backward MAC can be manually specified or randomly selected by the device.

- When maintenance engineers need to measure the packet loss rate of a specific user's service flows on a link, use the **onward-mac** and **backward-mac** parameters to manually specify the MAC address of the user terminal.
- When maintenance engineers need to measure the packet loss rate of the whole links, use the **update-onward-mac** and **update-backward-mac** parameters to make the device randomly select the service flows from a user terminal as the test sample.

Assume that, in **Figure 5-38**, the priority of the upstream and downstream packets of user 1's and user 2's HSI services is 0; the MEP on the OLT functions as an originating end; the MEP on the MDU functions as a responding end. The following table lists PM instance configurations (using **Figure 5-38** as an example).

Test Object	PM Instance Configuration on the MDU	PM Instance Configuration on the OLT
Data services of user 1	**pm instance mep 0/0/2 remote-mep 1 receive-8021p 0 send-8021p 0 backward-mac 00e0-fc00-1010**	**pm instance mep 0/0/1 remote-mep 2 receive-8021p 0 send-8021p 0 onward-mac 00e0-fc00-1010**
HSI services of user 1	**pm instance mep 0/0/2 remote-mep 1 receive-8021p 0 send-8021p 0 update-backward-mac**	**pm instance mep 0/0/1 remote-mep 2 receive-8021p 0 send-8021p 0 update-onward-mac**
HSI services of user 2	**pm instance mep 0/0/3 remote-mep 1 receive-8021p 0 send-8021p 0 update-backward-mac**	**pm instance mep 0/0/1 remote-mep 3 receive-8021p 0 send-8021p 0 update-onward-mac**

On a P2MP connection network, the access device differentiates between service flows according to the MAC addresses of user terminals. The following requirements must be met during PM instance deployment.

- The MAC address learning function is not disabled at the access end (for example, the MDU in **Figure 5-38**).
- The maximum number of learnable dynamic MAC addresses on the service port must be greater than or equal to the actual number of MAC addresses on the service port to ensure that all the source MAC addresses of packets sent by user terminals can be learnt by the access device. Otherwise, the access device may regard the user packets as unknown unicast packets and discard them. Therefore, the test results may be inaccurate.
- On an FTTB/FTTC network, if an up MEP is configured on the MDU service board, 1:1 VMAC can be enabled on the MDU, while N:1 VMAC address cannot.

 Copyright © Huawei Technologies Co., Ltd.

- In an FTTH network, 1:1 VMAC or N:1 VMAC cannot be enabled on the OLT.
- The access device implements single-ended Ethernet loss measurement (ETH-LM) on a P2MP connection network according to Huawei's proprietary protocol, and therefore the access devices interconnect only with Huawei devices. When the peer MEP is on a device which is not manufactured by Huawei, ETH-SLM can be used to collect the packet loss rate on a P2MP connection network.
- If the performance of the service flow needs to be monitored, the traffic classification parameter of the **service-port** must be created based on user-vlan instead of **user-8021p**, **user-8021p-list** or **user-encap**.

5.3.6.4.2 Performance Monitoring with Typical QoS Planning

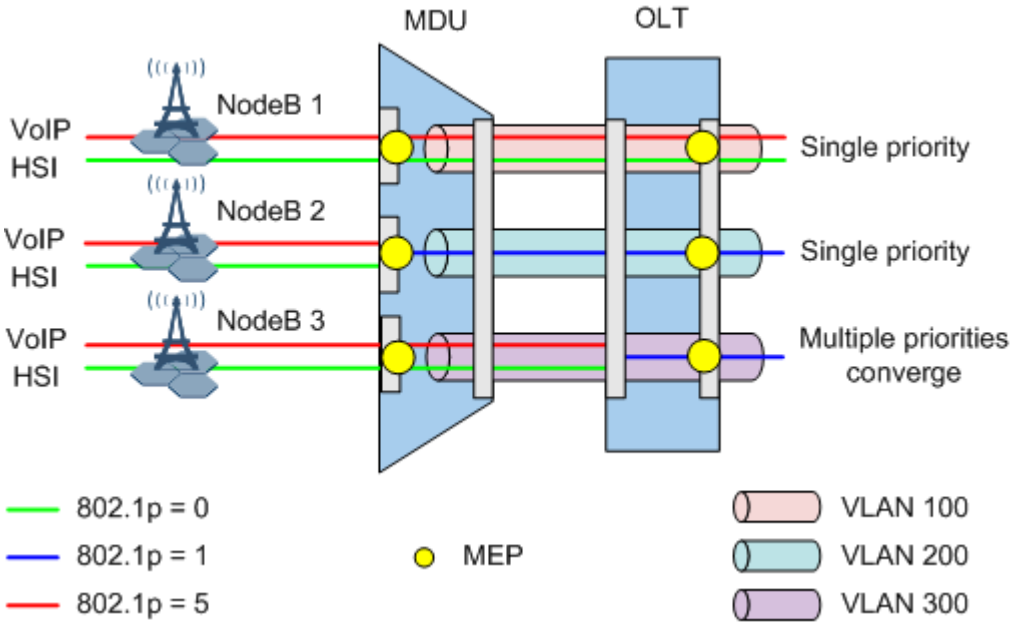
As defined in ITU-T Y.1731, the priority of Ethernet protocol packets used for single-ended Ethernet loss measurement (ETH-LM), two-way Ethernet delay measurement (ETH-DM) and single-ended Ethernet synthetic loss measurement (ETH-SLM) must be configurable to make the priority of Ethernet protocol packets equal to that of to-be-tested packets. If the priority of to-be-tested packets changes during packet forwarding, the priority of Ethernet protocol packets must be changed accordingly. The following describes the PM configuration and implementation principles on an FTTM point-to-point (P2P) network with different QoS plans.

Single Priority and Multi-priority Convergence

A service flow may contain packets with different priorities. Generally, the system chooses packets with the same priority for PM tests. This belongs to the single priority scenario (for example, NodeB 1 and NodeB 2 in **Figure 5-39**). However, in actual networking applications, carriers may modify packet priorities based on the QoS planning. Therefore, a service flow on a MEP may contain packets with different priorities, but this service flow on another MEP contains packets with the same priority. This belongs to the multi-priority convergence scenario (for example, NodeB 3 in the **Figure 5-39**).

 Copyright © Huawei Technologies Co., Ltd.

Figure 5-39 Single priority and multi-priority convergence



Assume that the 802.1p field is used to differentiate between service flows. Carriers configure different priority processing policies for different NodeBs, as shown in **Figure 5-39**.

- NodeB 1: The MDU and OLT copy the 802.1p priority of user-side packets as the priority of upstream packets and copy the 802.1p priority of network-side packets as the priority of downstream packets.
- NodeB 2: The MDU specifies the 802.1p priority of upstream packets to 1, and the OLT copies the 802.1p priority of user-side packets as the priority of upstream packets. The MDU and OLT copy the 802.1p priority of network-side packets as the priority of downstream packets.
- NodeB 3: The MDU copies the 802.1p priority of user-side packets as the priority of upstream packets, and the OLT specifies the 802.1p priority of upstream packets to 1. The MDU and OLT copy the 802.1p priority of network-side packets as the priority of downstream packets.

As shown in **Figure 5-39**, MEPs are deployed on the MDU user port and the OLT upstream port to monitor the performance of Ethernet links.

- For NodeB 1 and NodeB 2, MEPs at both ends of a link perform PM tests based on packets with the same priority in upstream and downstream directions. This belongs to the single priority scenario.
- For NodeB 3, a service flow on the MEP of the MDU contains packets with different priorities, while the service flow on the MEP of the OLT contains packets with the same priority. This belongs to the multi-priority convergence scenario. However, this scenario is not defined in ITU-T Y.1731. Therefore, the access device implements ETH-LM, ETH-DM and ETH-SLM of the multi-priority convergence scenario based on the Huawei's proprietary protocol. As defined in Huawei's proprietary protocol, the priority of Ethernet protocol packets on the MEP containing packets with different priorities is set to **unaware** so that all the packets sent to or received by the MEP are counted during statistics measurement. For example, to perform LM between the user

 Copyright © Huawei Technologies Co., Ltd.

port of the MDU and the upstream port of the OLT, set the priority of Ethernet protocol packets of the PM instance corresponding to NodeB 3 to **unaware** on the MDU.

Assume that the 802.1p priority of downstream packets is set to 0; the MEP on the OLT functions as an originating end; the MEP on the MDU functions as a responding end. The following table lists the priority configurations of Ethernet packets in PM instances. The priorities of services on NodeB 2 and NodeB 3 are modified to the same priority. Therefore, the system can monitor only the performance of all the services, but cannot monitor the performance of a specific service.

Test Object	Priority Configurations on the MDU	Priority Configurations on the OLT
VoIP services of NodeB 1	**receive-8021p 0 send-8021p 5**	**receive-8021p 5 send-8021p 0**
HSI services of NodeB 1	**receive-8021p 0 send-8021p 0**	**receive-8021p 0 send-8021p 0**
Services of NodeB 2	**receive-8021p 0 send-8021p 1**	**receive-8021p 1 send-8021p 0**
Services of NodeB 3	**receive-8021p 0 send-8021p unaware**	**receive-8021p 1 send-8021p 0**

In the multi-priority convergence scenario, the priority after multiple priorities converge is used as the priority of Ethernet protocol packets, which is different from the priority of the service flow. Therefore, PM results may be incorrect.

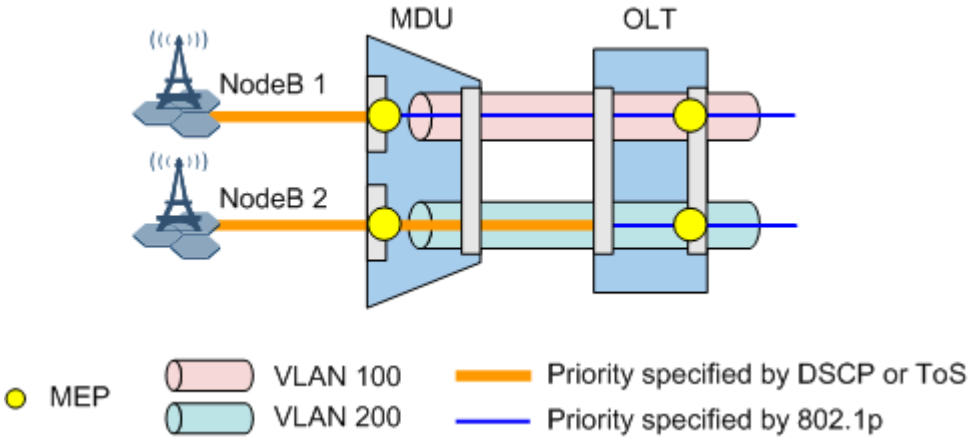
For example, if PM is performed on NodeB 3 in **Figure 5-39**, service packets forwarded by the MDU to the upstream direction are with different priorities, and the 802.1p priority of Ethernet protocol packets sent by the MEP on the MDU is 1. The priority of the Ethernet protocol packets is not the same as that of any service packets. In this case, the sequence of Ethernet protocol packets and service packets received on the OLT may change, which may be different from that of Ethernet protocol packets and service packets sent by the MDU. Consequently, the PM results may be inaccurate.

In the multi-priority convergence scenario, to ensure accurate PM results, ensure that the originating end, responding end, and devices in between map packets with different priorities into the same queue for scheduling so that the sequence of Ethernet protocol packets and service packets is not changed during scheduling.

Copyright © Huawei Technologies Co., Ltd.

Setting the 802.1p Priority Based on the ToS and DSCP Priorities

Figure 5-40 Setting the 802.1p priority based on the ToS and DSCP priorities



Assume that the ToS or DSCP field is used to differentiate between service flows on NodeBs, carriers configure different priority processing policies for different NodeBs, as shown in **Figure 5-40**.

- NodeB 1: The MDU sets the 802.1p priority of upstream packets based on the ToS or DSCP priority of the IP packets, and the OLT copies the 802.1p priority of user-side packets as the priority of upstream packets. The MDU and OLT copy the 802.1p priority of network-side packets as the priority of downstream packets.
- NodeB 2: The MDU forwards upstream packets based on the priority of the IP packets, and the OLT sets the 802.1p priority of the upstream packets based on the ToS or DSCP priority of the IP packets. The MDU and OLT copy the 802.1p priority of network-side packets as the priority of downstream packets.

As shown in **Figure 5-40**, MEPs are deployed on the MDU user port and the OLT upstream port to monitor the performance of Ethernet links.

- For NodeB 1, MEPs at both ends of a link perform PM tests based on packets with the same 802.1p priority in upstream and downstream directions (the 802.1p priority is the one that the MDU sets based on the ToS or DSCP priority). This belongs to the single-priority scenario.
- For NodeB 2, the MEP on the MDU performs PM tests based on the packets with the ToS or DSCP priority in upstream and downstream directions. The priority of the Ethernet protocol packets sent by the MDU is the same as the ToS or DSCP priority of service packets. The priority of service packets received on the OLT is the 802.1p priority that the OLT sets based on the ToS and DSCP priority. However, Ethernet protocol packets are non-IP packets and they do not carry the ToS or DSCP field. When the OLT modifies the packet priority, the OLT cannot set the priority of Ethernet protocol packets to be the same as that of service packets. In this case, packet disorder occurs among Ethernet protocol packets and service packets, and consequently the PM results are incorrect. Therefore, PM is not supported in this case.

Assume that the 802.1p priority of upstream service packets is set to 1 and that of downstream service packets is set to 0 based on the ToS or DSCP priority of IP packets; the MEP on the OLT functions as an originating end; the MEP on the

 Copyright © Huawei Technologies Co., Ltd.

MDU functions as a responding end. The following table lists the priority configurations of Ethernet packets in PM instances.

Test Object	Priority Configurations on the MDU	Priority Configurations on the OLT
Services of NodeB 1	receive-8021p 0 send-8021p 1	receive-8021p 1 send-8021p 0
Services of NodeB 2	Not supported	Not supported

5.3.6.5 PM Configuration

This topic describes how to configure ITU-T Y.1731-defined performance monitoring (PM) using the CLI. The CLI supports only the configuration of on-demand PM tests, such as Ethernet loss measurement (ETH-LM), Ethernet delay measurement (ETH-DM) and Ethernet synthetic loss measurement (ETH-SLM). Periodic PM tests are configured only on the NMS. For the detailed operations, see the NMS configuration guide.

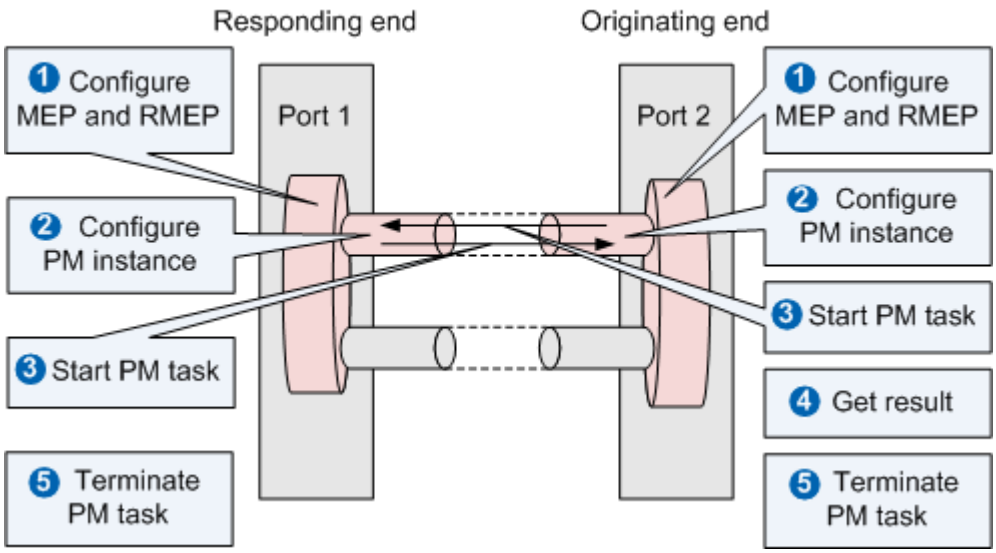
Prerequisites

Service configurations (such as VLAN and service flow configurations) are completed, and user-side services run normally.

Configuration Procedure

Figure 5-41 shows the overall procedure of configuring and performing a PM task on the originating and responding ends.

Figure 5-41 PM configuration procedure



 Copyright © Huawei Technologies Co., Ltd.

As shown in **Figure 5-41**, the configuration items and steps on the originating end are similar to those on the responding end. The following table lists the related configuration commands and precautions.

Step	Command	Precautions
1. Configure maintenance association end points (MEPs).	**cfm md**	Parameter **name-format** must be set to **no-name**.
	cfm ma	The maintenance association (MA) ID must be the same as the maintenance entity group (MEG) ID, and parameter **name-format** must be set to **icc-based**.
	cfm ma vlan	If the MEP is configured on the NXED board, the VLAN associated with the MA must be the same as that configured in **vlantag1** of the MEP in the MA. **NOTE** For a MEP on the ONT, do not configure this command.
	cfm ma meplist	The MEP list must contain the MEP and RMEP. **NOTE** For a MEP on the ONT, do not configure this command.
	cfm mep	For detailed limitations on MEP configurations, see Feature Dependency and Limitation.
2. Configure PM instances.	**pm instance**	• The parameters that need to be configured varies with the VLAN and QoS planning. For details, see **5.3.6.4.1 Performance Monitoring with Typical VLAN Planning** and **5.3.6.4.2 Performance Monitoring with Typical QoS Planning**. • You are advised to set the test ID on the originating end to be the same as that on the responding end to facilitate future management and maintenance if these 2 test IDs belong to the same PM instance. • The **display pm instance** command can be used to query the configured PM instances in the system.

Copyright © Huawei Technologies Co., Ltd.

Step	Command	Precautions
3. Start PM tasks.	ETH-LM 1. Responding end: **pm lm receive** 2. Originating end: **pm lm send** ETH-DM 1. Responding end: **pm dm receive** 2. Originating end: **pm dm send** ETH-SLM 1. Responding end: **pm slm receive**	• The number of started PM tasks must comply with that of PM tasks that can be supported by the devices where the originating and responding ends are configured. • For an LM test, if one of the 2 MEPs (belonging to the same PM instance) is set as the originating end on a device, the other MEP cannot be set as the responding end on the same device. Similarly, if one of the 2 MEPs is set as the responding end on a device, the other MEP cannot be set as the originating end on the same device. • After a PM task is started on the responding end, the responding end sends the response packets to the peer MEP only when the test ID configured for the PM task on the responding end matches the MEP configured on the originating end. • The **display pm task** command can be used to query the started PM tasks on the device. If a PM task is interrupted abnormally, but the timeout interval for the PM task does not expire, this PM task can be queried by running the **display pm task** command. • The **display pm oam pdu statistics** command can be used to query the Ethernet protocol packets that are transmitted and received by the device.
4. Obtain the test results.	**display pm statistics**	You can use any of the following methods to obtain the test results on the originating end. • The test results will be displayed on the CLI in real time after a PM task is started. • You can run the **display pm statistics** command to query the test results after the test is completed.

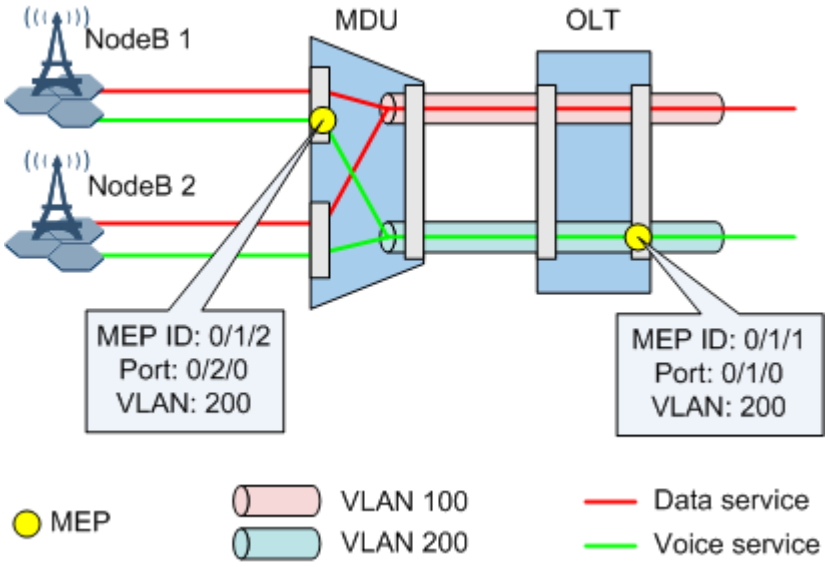
 Copyright © Huawei Technologies Co., Ltd.

Step	Command	Precautions
5. Stop a PM task.	ETH-LM 1. Originating end: - 2. Responding end: **undo pm lm receive** ETH-DM 1. Originating end: - 2. Responding end: **undo pm dm receive** ETH-SLM 1. Responding end: **undo pm slm receive**	A PM task on the originating end will be stopped if any of the following conditions is met. • The maintenance engineer presses **Ctrl+C** to manually stop the PM task. • The PM task is completed successfully, or the timeout interval for a PM task has expired so that the system automatically stops the PM task.

PM Instance and Data Planning

The following describes how to configure PM on a fiber to the mobile base station (FTTM) network.

Figure 5-42 LM on an FTTM network



Assume that mobile services in **Figure 5-42** are transmitted upstream through the NXED board on the OLT. If maintenance engineers need to perform an LM test on

Copyright © Huawei Technologies Co., Ltd.

NodeB 1's voice services on an FTTM network, deploy the MEPs on the MDU user port and the OLT upstream port as shown in **Figure 5-42**. The MEP on the OLT functions as the originating end, and the MEP on the MDU functions as the responding end. The following table lists the data planning.

Item	Data
MEP	The MEP ID and the port where the MEP is configured are shown in **Figure 5-42**. MEP level: 0 VLAN planning of the MEP: N:1 VLAN (as listed in the following) • Customer VLANs (C-VLANs) are used to differentiate between upstream service flows of a NodeB. • The MDU translates C-VLANs to service VLANs (S-VLANs). Packets of the same type are transmitted upstream through the same S-VLAN. S-VLAN 100 is intended for data services, while S-VLAN 200 is intended for voice services. • The OLT transparently transmits VLANs.
PM instance	Test ID planning: • Data services: test ID 100 • Voice services: test ID 200 QoS planning: The MDU and OLT copy the priority of user-side packets as the priority of upstream packets, and copy the priority of network-side packets as the priority of downstream packets. • Data services: The priority of upstream packets is 0 and that of downstream packets is 1. • Voice services: The priority of both upstream and downstream packets is 5. Onward MAC and backward MAC: The device randomly chooses the source MAC addresses of the service packets sent from a NodeB as the onward and backward MAC addresses.
PM task	The Ethernet protocol packets used for an LM test are sent every 1s and for 4 times.

Procedure

Step 1 Configure MEPs.

MDU configurations:

```
MDU (config) #cfm md 0 name-format no-name level 0
MDU (config) #cfm ma 0/1 name-format icc-based FTTM_VOICE
MDU (config) #cfm ma 0/1 vlan 200
MDU (config) #cfm ma 0/1 meplist 1
MDU (config) #cfm ma 0/1 meplist 2
MDU (config) #cfm mep 0/1/2 direction up port 0/2/0 vlantag1 200
```

 Copyright © Huawei Technologies Co., Ltd.

OLT configurations:

```
OLT(config)#cfm md 0 name-format no-name level 0
OLT(config)#cfm ma 0/1 name-format icc-based FTTM_VOICE
OLT(config)#cfm ma 0/1 vlan 200
OLT(config)#cfm ma 0/1 meplist 1
OLT(config)#cfm ma 0/1 meplist 2
OLT(config)#cfm mep 0/1/1 direction up port 0/1/0 vlantag1 200
```

Step 2 Configure PM instances.

MDU configurations:

```
MDU(config)#pm instance test-id 200 mep 0/1/2 remote-mep 1 receive-8021p 5 send-8021p 5 update-backward-mac
```

OLT configurations:

```
OLT(config)#pm instance test-id 200 mep 0/1/1 remote-mep 2 receive-8021p 5 send-8021p 5 update-onward-mac
```

Step 3 Start a PM task and obtain the test results.

MDU configurations:

```
MDU(config)#pm lm receive single-ended test-id 200
```

OLT configurations:

```
OLT(config)#pm lm send single-ended test-id 200 interval 1000 count 4

 Command:
        pm lm send single-ended test-id 100 interval 1000 count 4
 Press CTRL_C to break
 ---------------------------------------------------------------------------
 Index    Near-loss    Near-loss     Far-loss     Far-loss    ERR
                  ratio                    ratio
 ---------------------------------------------------------------------------
    1          0          0          8      7407407      1
    2          0          0        100      1000000      0
    3       8000   80000000        900     90000000      0
 ---------------------------------------------------------------------------
 Maximum near-loss : 8000       Maximum near-loss ratio : 80000000
 Minimum near-loss : 0          Minimum near-loss ratio : 0
 Average near-loss : 2666       Average near-loss ratio : -
 Total near-loss  : -
 Maximum far-loss : 900         Maximum far-loss ratio : 90000000
 Minimum far-loss : 8           Minimum far-loss ratio : 1000000
 Average far-loss : 336         Average far-loss ratio : 9074540
 Total far-loss : 1008
 Note:
    Ratio unit: 10^(-8)
    ERR:
      0: No error
      1: Statistics for single-ended LM are incorrect
      2: Statistics for average single-ended LM ratio are incorrect
      3: Statistics for two-way DM are incorrect
      4: A value exceeds the threshold and therefore a reverse occurs
```

Step 4 Stop a PM task.

In this test, the PM task is completed on the originating end. Maintenance engineers only need to run the following command to stop the PM task on the responding end. MDU configurations:

```
MDU(config)#undo pm lm receive single-ended test-id 200
```

----End

 Copyright © Huawei Technologies Co., Ltd.

Result

The LM test result (including the number of lost packets on the local and remote ends and packet loss rate) will be displayed on the CLI, as listed in step 3. ITU-T Y.1731 provides the detailed explanation about the parameters in the test results.

Follow-up Procedure

Maintenance engineers diagnose whether the link between the 2 MEPs is faulty based on the test results.

 Copyright © Huawei Technologies Co., Ltd.

6 MPLS

Multiprotocol Label Switching (MPLS) was introduced to improve the forwarding speed. However, because of its excellent performance in traffic engineering (TE) and virtual private network (VPN), which are the two critical technologies, MPLS is becoming an important standard for extending the IP network.

6.1 Overview

6.2 MPLS Reference Standards and Protocols

6.3 MPLS Overview
Multiprotocol Label Switching (MPLS) was introduced to improve the forwarding speed. However, because of its excellent performance in traffic engineering (TE) and virtual private network (VPN), which are the two critical technologies, MPLS is becoming an important standard for extending the IP network. This topic provides the introduction, availability, principle, and reference of the MPLS feature.

6.4 MPLS RSVP-TE
MPLS RSVP-TE is a technology which integrates TE and the MPLS superimposed model. It provides high quality of service (QoS) and TE capability for users by establishing LSPs based on TE. This topic provides introduction to this feature and describes the principle and reference documents of this feature.

6.5 MPLS OAM
MPLS OAM checks if an LSP is in the normal state through a mechanism, and reports the alarm information if the LSP fails. This topic provides introduction to this feature and describes the principle and reference documents of this feature.

6.6 MPLS TE Reliability

6.7 Configuring the MPLS Service
This topic describes the MPLS technology and how to configure the MPLS service on the MA5800.

6.1 Overview

Multi-protocol Label Switching (MPLS) is between the data link layer and the network layer in the TCP/IP protocol stack. The label in a short fixed length is used to encapsulate IP packets. On the data plane, fast label forwarding is implemented. On the control plane, MPLS can meet the requirements on the

 Copyright © Huawei Technologies Co., Ltd.

network from various new applications with the help of the powerful and flexible routing functions of the IP network.

The MPLS feature includes the following sub features:

- Basic MPLS functions

  Basic MPLS functions provide a basis for other MPLS sub features. MPLS, which is not restricted by any specific link layer protocol, can use any Layer 2 medium to transmit network packets. This shows that MPLS is not a service or application, but a tunnel technology. This technology can support multiple higher-layer protocols and services, and ensure the security of information transmission at the same time.
- MPLS RSVP-TE

  To deploy engineered traffic on a large-scale backbone network, a simple solution with good expansibility must be adopted. MPLS, as a stacking model, can easily establish a virtual topology over a physical network and map traffic to this topology. Therefore, a technology that integrates MPLS with traffic engineering, namely, MPLS-TE is generated.
- MPLS OAM

  MPLS, as the key bearer technology for the extensible network-generation network, provides multiple services with QoS guarantee. In addition, MPLS introduces a unique network layer and therefore the faults caused by this new network layer may occur. Therefore, an MPLS network must have the OAM capability.

The MPLS feature supports the following functions:

- Functioning as a P device
- Capability of 100 pps for processing LDP and RSVP packets when functioning as a P device
- MPLS label switching
- Penultimate hop popping (PHP)
- Query of LSP packet statistics by label

6.2 MPLS Reference Standards and Protocols

The following lists the reference standards and protocols of this feature:

1. PWE3
   - RFC3985: Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture
   - RFC4447: Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)
   - RFC3916: Requirements for Pseudo-Wire Emulation Edge-to-Edge (PWE3)
   - RFC4446: IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)
   - RFC4717: Encapsulation Methods for Transport of Asynchronous Transfer Mode (ATM) over MPLS Networks
   - RFC4448: Encapsulation Methods for Transport of Ethernet over MPLS Networks

 Copyright © Huawei Technologies Co., Ltd.

- RFC5085: Pseudowire Virtual Circuit Connectivity Verification (VCCV): A Control Channel for Pseudowires
- RFC4553: Structure-Agnostic Time Division Multiplexing (TDM) over Packet (SAToP)
- RFC5462: Multiprotocol Label Switching (MPLS) Label Stack Entry: EXP Field Renamed to Traffic Class Field
- RFC4385: Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN
- RFC6870: Pseudowire Preferential Forwarding Status Bit

2. RSVP
   - RFC2205: Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification
   - RFC3209: RSVP-TE: Extensions to RSVP for LSP Tunnels
   - RFC2210: The Use of RSVP with IETF Integrated Services
   - RFC2961: RSVP Refresh Overhead Reduction Extensions
   - RFC3270: Multi-Protocol Label Switching (MPLS) Support of Differentiated Services
   - RFC4090: Fast Reroute Extensions to RSVP-TE for LSP Tunnels
3. LDP
   - RFC3031: Multiprotocol Label Switching Architecture
   - RFC5036: LDP Specification
   - RFC3215: LDP State Machine
   - RFC3478: Graceful Restart Mechanism for Label Distribution Protocol
   - RFC3815: Definitions of Managed Objects for the Multiprotocol Label Switching (MPLS), Label Distribution Protocol (LDP)
4. MPLS
   - draft-ietf-mpls-lsp-ping-version-06
   - RFC4379: Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures
   - RFC3032: MPLS Label Stack Encoding
   - RFC3469: Framework for Multi-Protocol Label Switching (MPLS)-based Recovery
   - RFC3812: Multiprotocol Label Switching (MPLS) Traffic Engineering (TE) Management Information Base (MIB)
   - RFC3813: Multiprotocol Label Switching (MPLS) Label Switching Router (LSR) Management Information Base (MIB)
   - RFC3814: Multiprotocol Label Switching (MPLS) Forwarding Equivalence Class To Next Hop Label Forwarding Entry (FEC-To-NHLFE) Management Information Base (MIB)
   - Y.1710: Requirements for OAM functionality for MPLS networks
   - Y.1711: OAM mechanisms for MPLS networks
   - Y.1720: Protection switching for MPLS networks

 Copyright © Huawei Technologies Co., Ltd.

6.3 MPLS Overview

Multiprotocol Label Switching (MPLS) was introduced to improve the forwarding speed. However, because of its excellent performance in traffic engineering (TE) and virtual private network (VPN), which are the two critical technologies, MPLS is becoming an important standard for extending the IP network. This topic provides the introduction, availability, principle, and reference of the MPLS feature.

6.3.1 Introduction

Definition

Basic MPLS features mainly refer to the MPLS Label Distribution Protocol (LDP) and LSP management function.

The LDP protocol is a standard MPLS label distribution protocol defined by the IETF. LDP, which is mainly used to allocate labels for the negotiation between LSRs to set up label switching paths (LSPs), regulates various types of information for the label distribution process, and the related processing. The LSRs form an LSP that crosses the entire MPLS domain according to the local forwarding table, which correlates in the label, network hop node, and out label of each specific FEC.

With the LSP management function, the MA5800 can manage and maintain the LSPs generated by various LDPs and can issue the hardware forwarding module.

Purpose

MPLS is initially put forth to improve the forwarding speed of routers. Compared with the traditional IP routing mode, during data forwarding, MPLS analyzes the IP packet header only on the edge of the network, but does not analyze the IP packet header at each hop. This saves the processing time.

With the development of the ASIC technology, the route search speed is not a bottleneck for network development. Thus, MPLS has not obvious advantages in forwarding speed. MPLS, however, is widely applied to the virtual private network (VPN), traffic engineering, and quality of service (QoS) due to its characteristics of supporting multi-layer labels and connected-oriented forwarding plane. Therefore, MPLS becomes an increasingly important standard for expanding the scale of the IP network.

6.3.2 Principle

Multiprotocol label switching (MPLS) was introduced to improve the forwarding speed. However, because of its superb performance in traffic engineering (TE) and virtual private network (VPN), which are the two critical technologies in the current IP network, MPLS has become an important standard for extending the IP network.

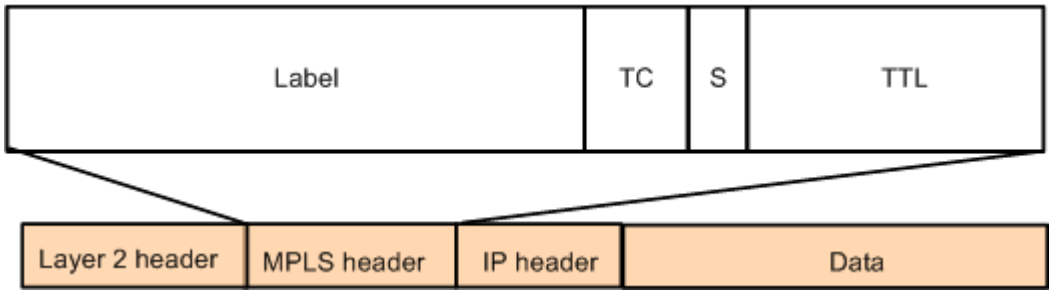
IP technologies are connectionless at both the forwarding plane and control plane whereas ATM technologies are connection-oriented at the two planes. The MPLS technology combines the advantages of IP and ATM technologies and achieves a

 Copyright © Huawei Technologies Co., Ltd.

connectionless control plane and a connection-oriented forwarding plane. Such a combination provides for flexible IP routing and convenient Layer 2 switching as well as expanded ATM service variety.

Figure 6-1 shows the MPLS packet format.

Figure 6-1 MPLS packet format



- Label: a 20-bit label value field, used as the forwarding pointer.
- TC: short for traffic class, a 3-bit field for QoS (note that this field was named EXP and is renamed TC in RFC5462).
- S: a 1-bit bottom of stack field. This bit set to 1 indicates the bottom label in the label stack.
- TTL: short for time to live, an 8-bit field, similar to the TTL field in an IP header.

Basic MPLS Concepts

- Forwarding equivalence class (FEC)

  An FEC refers to a group of data streams which are forwarded in the same manner. These data streams are forwarded by the LSR in the same manner. Theoretically, FECs can be classified according to the IP address, service type, or QoS. For example, in the conventional IP forwarding by using the maximum matching algorithm, all the packets to the same route belong to an FEC. Currently, FECs are generally classified based on the address. The MA5800 supports only address-based FECs.

- Label

  A label is a short fixed length physically contiguous identifier which is used to identify an FEC, usually of local significance. In certain conditions, for example, when load sharing is required, one FEC may map multiple labels. On one device, however, one label can represent only one FEC.

  Label encapsulation is performed between the link layer and the network layer. Therefore, label can be supported by any link layer.

- Penultimate hop popping

  On the last hop node, the label no longer has any function. In this case, the label stack may be popped at the penultimate LSR of the LSP, rather than at the LSP Egress, to reduce the load of the last hop LSR. The last hop LSR directly forwards IP packets or next-layer labels, which are configured at the egress by the PHP.

- Label switching router (LSR)

 Copyright © Huawei Technologies Co., Ltd.

An LSR, also called an MPLS node, is a network device which is capable of exchanging and forwarding MPLS labels. LSRs are the basic elements in an MPLS network. All LSRs support the MPLS protocol.

- Label edge router (LER)

  An LSR on the edge of the MPLS domain is called the LER. If an LSR has a neighbor node that does not run the MPLS protocol, the LSR is an LER.

  The LER is responsible for classifying the packets that enter the MPLS domain to FECs and adding labels to these FECs for forwarding in the MPLS domain. When the packets leave the MPLS domain, the FECs pop up the labels, resume the original packets, and then are forwarded accordingly.

- Label switched path (LSP)

  The path that a packet in a particular FEC traverses in an MPLS network is called the LSP.

  The LSP, similar to the ATM virtual circuit in function, is a unidirectional path from the ingress to the egress.

- Label distribution protocol (LDP)

  LDP, also called the signaling protocol, is the MPLS control protocol. LDP is responsible for series of operations such as FEC classification, label distribution, and LSP establishment and maintenance.

  MPLS can use multiple label distribution protocols, such as the Label Distribution Protocol (LDP) and Resource Reservation Protocol Traffic Engineering (RSVP-TE).

  – LDP is a standard MPLS label distribution protocol defined by the IETF. LDP is responsible for FEC classification, label distribution, and LSP establishment and maintenance.
  – RSVP-TE is an extension to RSVP and provides high QoS and TE capability for users by establishing TE LSPs.

- Label distribution mode

  In an MPLS system, the downstream LSR determines the label to be advertised to a specific FEC, and then notifies the upstream LSR. That is, the label is specified by the downstream LSR, and is advertised from the downstream LSR to the upstream LSR.

  The label advertisement modes on the upstream and downstream LSRs with label advertisement adjacencies must be the same. Otherwise, the LSP cannot be set up.

  The two label advertisement modes are as follows:

  – Downstream unsolicited (DU) mode

    In the DU mode, the LSR allocates labels to a specific FEC without asking for the label request message from upstream LSRs.

  – Downstream on demand mode

    In the DoD mode, the LSR allocates labels to a specific FEC only after obtaining the label request message from upstream LSRs.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

When a downstream LSR feeds back the label mapping information is determined by the label control mode used by the LSR.

- When an LSR supports the ordered label control mode, it sends the label mapping information to the upstream LSR only when it receives the label mapping message returned by the downstream LSR, or when it is the egress node of the FEC.
- When an LSR supports the independent label distribution control mode, it sends the label mapping message to the upstream LSR regardless of whether it receives the label mapping message returned by the downstream LSR.

- Label distribution control mode

  The label distribution control mode is the mode used by the LSR to allocate labels during the establishment of LSPs.

  The two label distribution control modes are as follows:

  – Independent label distribution control mode

    In the independent label distribution control mode, the local LSR can independently allocate a label to an FEC and binds the label to the FEC, and notify the upstream LSR of the label, without waiting for the label from the upstream LSR.

  – Ordered label control mode

    In the ordered label control mode, the LSR can send the label mapping message of an FEC to the upstream LSR only when the LSR has the label mapping message of the next hop of the FEC, or when the LSR is the egress node of the FEC.

- Label retention mode

  The label retention mode is the mode adopted by the LSR to process the received label mapping messages that are not in use temporarily.

  The two label retention modes are as follows:

  – Liberal retention mode

    If an LSR supports the liberal retention mode, it maintains the label mapping received from the neighbor LSR regardless of whether the neighbor LSR is its own next hop.

    When the next hop neighbor changes due to the change of network topology, the LSR that supports the liberal retention mode can use the label sent from the non-next-hop neighbor to set up LSPs quickly. This, however, requires more memory and label space.

  – Conservative retention mode

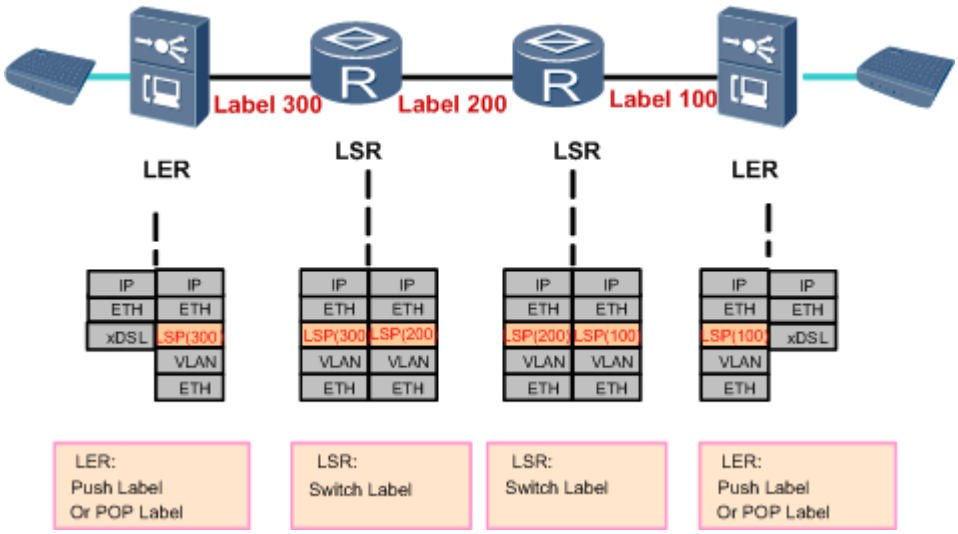
    If an LSR supports the conservative retention mode, it maintains the label mapping received from the neighbor LSR only when the neighbor LSR is its next hop.

    When the next hop neighbor changes due to the change of network topology, the LSR that supports the conservative retention mode can save memory and label space because the LSR maintains only the label from the next hop neighbor. The re-establishment of LSPs, however, lasts a long time.

Figure 6-2 shows the protocol stack model for label distribution.

 Copyright © Huawei Technologies Co., Ltd.

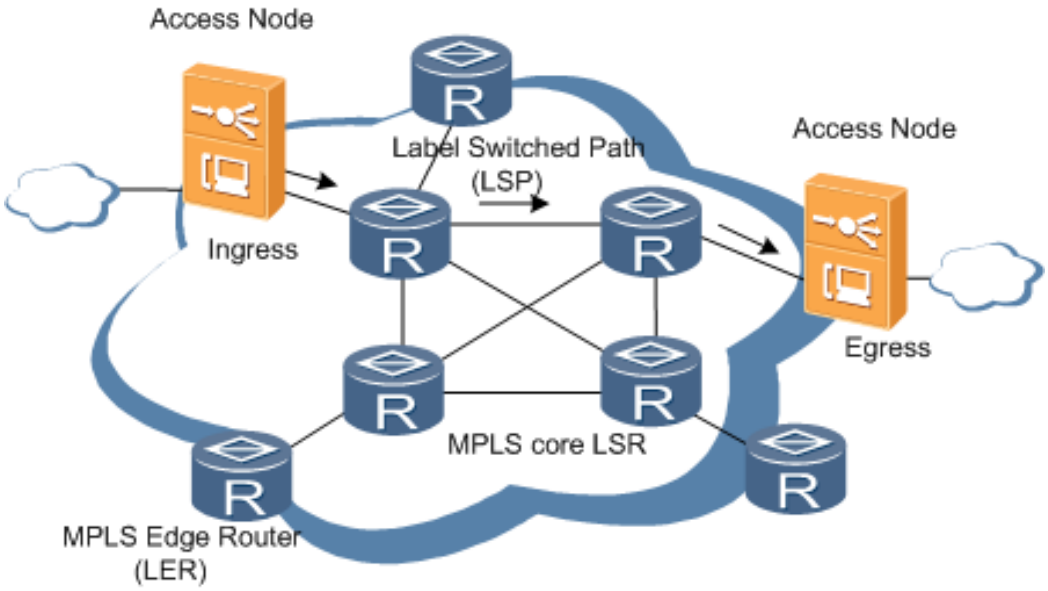
Figure 6-2 Protocol stack model for label distribution



Working principle of the MPLS feature

Figure 6-3 shows the working principle of the MPLS feature

Figure 6-3 MPLS network structure



1. First, enable MPLS and LDP on each router on the network, and enable LDP on the interconnected interfaces.
2. Consequently, LDP automatically sets up an LDP session between any two routers. The LDP packets are carried on this session.
3. LDP works with the traditional routing protocol such as OSPF and RIP to set LSPs in each LSR for the FEC with service requirements.

 Copyright © Huawei Technologies Co., Ltd.

4. LDP does not need to be enabled for the establishment of static LSPs. Configure the FEC, and inbound and outbound labels on each MPLS router that the static LSP travels.

LDP GR

The GR is a key technology for implementing the high availability (HA). The GR protocol collects the information about the protocol control plane from neighbors or remote peers but does not learn about the information about the control plane through the handshake and exchange of the protocol.

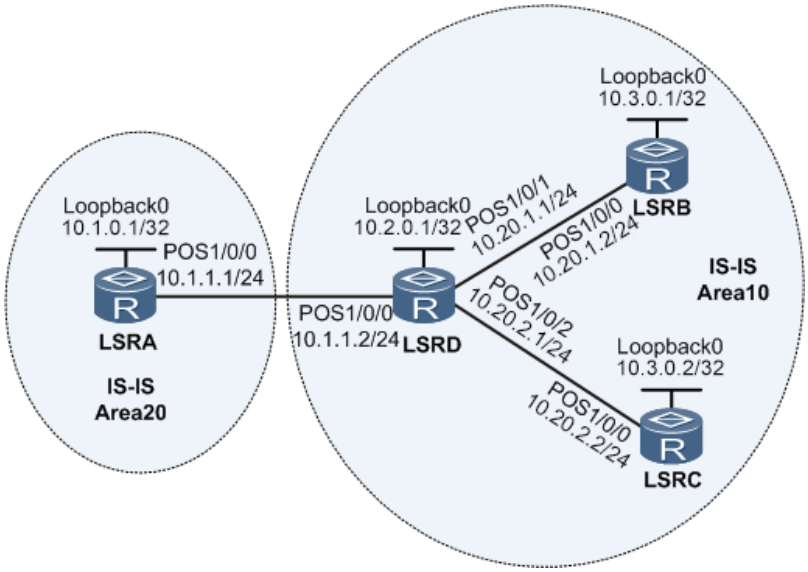
The LDP GR function ensures normal forwarding of the MPLS service during the active/standby switchover or upgrade of the system. In addition, the LDP GR function resumes the LDP session and completes the LSP establishment after the active/standby switchover or upgrade of the system

NOTE

In actual application, to prevent services from being affected by the active control board failure, configure the system-level GR in the environment where both active and standby control boards are configured.

LDP Extension for Inter-Area LSP

Figure 6-4 Networking topology of LDP Extension for Inter-Area LSP



As shown in **Figure 6-4**, there are two IGP areas, Area 10 and Area 20.

In the routing table of LSRD at the edge of Area 10, there are two host routes to LSRB and LSRC. Generally, to prevent a large number of routes from occupying too many resources, on LSRD, you can use IS-IS to aggregate the two routes to one route 10.3.0.0/24 and send this route to Area 20. Consequently, there is only one aggregated route (10.3.0.0/24) but not 32-bit host routes in the routing table of LSRA. By default, when establishing LSPs, LDP searches the routing table for the route that exactly matches the forwarding equivalence class (FEC) in the received Label Mapping message. **Table 6-1** shows routing entry information of LSRA and routing information carried in FEC in the situation as shown in **Figure 6-4**.

 Copyright © Huawei Technologies Co., Ltd.

Table 6-1 Routing entry information of LSRA and routing information carried in FEC

Routing entry information of LSRA	FEC
10.3.0.0/24	10.3.0.1/32
	10.3.0.2/32

LDP establishes liberal LSPs rather than inter-area LDP LSPs for aggregated routes. In this situation, LDP cannot provide required backbone network tunnels for VPN services.

Therefore, in the situation as shown in **Figure 6-4**, you need to configure LDP to search for routes according to the longest match rule to establish LSPs. There is already an aggregated route 10.3.0.0/24 in the routing table of LSRA. When LSRA receives a Label Mapping message (such as the carried FEC is 10.3.0.1/32) from Area 10, LSRA searches for a route according to the longest match rule defined in RFC 5283. Then, LSRA finds information about the aggregated route 10.3.0.0/24, and uses the outbound interface and next hop of this route as those of the route 10.3.0.1/32. In this manner, LDP can establish inter-area LDP LSPs.

6.4 MPLS RSVP-TE

MPLS RSVP-TE is a technology which integrates TE and the MPLS superimposed model. It provides high quality of service (QoS) and TE capability for users by establishing LSPs based on TE. This topic provides introduction to this feature and describes the principle and reference documents of this feature.

6.4.1 Introduction

Definition

MPLS RSVP-TE is a technology that integrates TE with the MPLS technology. MPLS RSVP-TE establishes label switched path (LSP) tunnels along specified paths for resource reservation, enables network traffic to avoid the node where congestion occurs to balance network traffic.

To establish constraint-based LSPs in MPLS TE, RSVP is extended. The extended RSVP signaling protocol is called the RSVP-TE signaling protocol.

Purpose

To deploy engineered traffic on a large-scale backbone network, a simple solution with good expansibility must be adopted. MPLS, as a stacking model, can easily establish a virtual topology over a physical network and map traffic to this topology.

MPLS TE establishes the LSP tunnel along a specified path through RSVP-TE and reserves resources. Thus, carriers can accurately control the path that traffic

 Copyright © Huawei Technologies Co., Ltd.

traverses to avoid the node where congestion occurs. This solves the problem that certain paths are overloaded and other paths are idle, utilizing the current bandwidth resources sufficiently. At the same time, MPLS TE can reserve resources during the establishment of LSP tunnels to ensure the QoS.

To ensure continuity of services, MPLS TE also introduces route backup to implement quick switching in case of link failure.

6.4.2 Principle

Basic MPLS RSVP-TE Concepts

- CR-LSP

  An LSP that is established based on certain constraints is called a constraint-based routed label switched path (CR-LSP). Different from a common LSP, the establishment of a CR-LSP depends on the routing information. In addition, some conditions must be met, for example, the specified bandwidth, the fixed route, and QoS parameters.

  CR-LSPs can be classified into the following two categories:

  – Static CR-LSP

    The forwarding information and resources information about a static CR-LSP are configured manually and the signaling protocol and route calculation are not involved. Less resource is occupied because the MPLS control packets do not need to be exchanged. The static CR-LSP, however, is seldom applied because it cannot dynamically adjust according to the topology change of the network.

  – Dynamic CR-LSP

    A dynamic CR-LSP is established and maintained through the signaling mechanism, and route calculation is required.

- RSVP

  Resource Reservation Protocol (RSVP) is designed for the integrated service model and is used to reserve resources on each node on a path. RSVP works on the transmission layer, but does not participate in the transmission of application data. RSVP, similar to ICMP, is a network control protocol.

- RSVP-TE

  To establish the CR-LSP, RSVP is extended. The extended RSVP signaling protocol is called the RSVP-TE signaling protocol.

- Explicit route

  A CR-LSP that is established along a specified path is called an explicit route. The two types of explicit route are as follows:

  – Strict explicit route

    On a strict explicit route, the next hop node must be directly connected to its preceding hop node. The route of the LSP can be precisely controlled by using the strict explicit route.

  – Loose explicit route

    The path between a loose node and its preceding node MAY include other network nodes that are not part of the strict node or its preceding abstract node.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

The MPLS TE signaling can carry the strict or loose attributes of an explicit path, and establish a CR-LSP along a specified path.

Composition of MPLS RSVP-TE

The following four components are necessary to the MPLS TE function:

- Information advertisement component

  In addition to the topology information about the network, TE also needs to know the load information about the network. Therefore, MPLS TE introduces the information advertisement component, that is, MPLS TE maintains the link attribute and topology attribute of the network on each node through IGP extensions to form the TE database (TEDB). The path that meets all types of constraints can be calculated by using the TEDB. The extended OSPF protocol adds certain TE-related attributes such as link bandwidth and color to the link connection status, where the maximum reservable bandwidth and unreserved bandwidth for the link with each priority are the most important.

- Route selection component

  After the information advertisement component forms the TEDB, the path that the LSP tunnel passes can be specified on each ingress node. This explicit path can be a strict or loose explicit path. In addition, the restraints such as the bandwidth can be specified.

  The route selection component calculates the path that meets the specified constraints by using the data in the TEDB through the constraint shortest path first (CSPF) algorithm.

- Signaling component

  After the shortest path from the ingress to the egress of the LSP is obtained, the TE tunnel, which is used to forward the traffic that enters the ingress of the LSP, needs to be established. This process is implemented by the signaling component.

  The MA5800 supports establishment of LSP tunnels through RSVP. The RSVP signaling can carry the constraint parameters such as the bandwidth of the LSP, certain explicit routes, and color.

  An LSP can also be established without the signaling protocol. That is, an LSP can be established through allocating labels manually hop by hop. An LSP established in this mode is called a static CR-LSP.

- Packet forwarding component

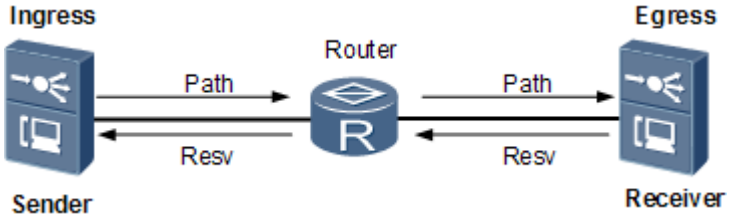
  The packet forwarding component of MPLS RSVP-TE is based on the label, that is, it forwards packets along the existing LSPs through labels. The defects of the IGP routing protocol can be avoided because the path of an LSP tunnel can be specified.

Process of TE LSP Tunnel Establishment

The LSP established through RSVP-TE has the resource reservation capability, and certain resources of the LSR on the LSP can be allocated to the LSP. Thus, the services transmitted on the LSP can be guaranteed. **Figure 6-5** shows the process of TE LSP tunnel establishment.

 Copyright © Huawei Technologies Co., Ltd.

Figure 6-5 Process of TE LSP tunnel establishment



The process of TE LSP tunnel establishment is summarized as follows:

1. The ingress LSR generates the Path message and transmits it to the egress LSR.
2. After the egress LSR receives the Path message, the egress LSR generates the Resv message and transmits it to the ingress LSR. At the same time, the LSRs on the LSP reserves resources for the LSP through the Resv message.
3. When the ingress LSR receives the Resv message, it indicates that the LSP is successfully established.

RSVP-TE GR

RSVP-TE graceful restart (GR) is a status recovery mechanism of RSVP-TE. When the control plane performs active/standby switchover, RSVP-TE GR can ensure the continuity of data transmission on the forwarding plane. At the same time, neighbor nodes help the GR node to recover in time.

RSVP-TE GR is based on the Hello mechanism of RSVP. The recovery of the local status depends on the upstream Path message or the downstream Recovery Path message.

RSVP GR has the following features: Shortening the information recovery of the control plane; reducing changes of temporary routes; ensuring the continuity of service forwarding on the forwarding plane.

6.5 MPLS OAM

MPLS OAM checks if an LSP is in the normal state through a mechanism, and reports the alarm information if the LSP fails. This topic provides introduction to this feature and describes the principle and reference documents of this feature.

6.5.1 Introduction

Definition

Operation Administration & Maintenance (OAM) has the following features:

- Simplifying network operations
- Checking the network performance anytime
- Reducing OPEX of the network

 Copyright © Huawei Technologies Co., Ltd.

Deployment of an effective OAM mechanism is crucial to the running of the network, especially to the network with certain QoS requirements, namely, certain performance and usability requirements.

MPLS, as the key bearer technology for the extensible network generation network, provides multiple services with QoS guarantee. In addition, MPLS introduces a unique network layer and therefore there will be faults that are only relevant to this new network layer. Therefore, an MPLS network must have the OAM capability.

MPLS OAM provides both detection tools and mature protection switching mechanisms. In this way, MPLS can perform switching when a fault occurs on the MPLS layer. This minimizes the loss of data.

Purpose

The MPLS OAM functions are as follows:

- Fault detection: Requirement-based query and continuous detection are provided to learn about anytime whether faults exist on the monitored LSP.
- Protection switching: After a fault occurs, it can be detected, analyzed, and located, and an alarm will be reported. In addition, the corresponding measures can be taken according to the fault type.

6.5.2 Principle

Background Knowledge for MPLS OAM

1. MPLS OAM packets are classified as follows:
   - Connectivity detection (CD) packets. The two types of CD packets are as follows:
     - Connectivity verification (CV)
     - Fast failure detection (FFD)
   - Forward defect indication (FDI)
   - Backward defect indication (BDI)

   MPLS OAM is implemented by periodically transmitting detection packets CV or FFD over the detected LSPs.
2. Basic detection process

   MPLS OAM is implemented by periodically transmitting detection packets CV and FFD over the detected LSPs.
   - To detect the source by using the CV packet, a sliding window in the width of 3s is set on the source and the LSP status is checked by using the VC packet received in the sliding window.
   - To detect the source by using the FFD packet, a sliding window in the width of three times of FFD transmit interval is set on the source and the LSP status is checked by using the FFD packet received in the sliding window.
3. CV and FFD

 Copyright © Huawei Technologies Co., Ltd.

The FFD and CV detection packets are mutually exclusive. That is, only the FFD or CV detection packets can be applied to one LSP at a time.

4. Backward path

   BDI packets are transmitted through the backward path. The ingress of a backward path is the egress of the detected LSP, and the egress of the backward path is the ingress of the detected LSP. That is, each forward LSP has a backward path.

5. Protection switching (PS)

   When a fault occurs on the network, currently MPLS OAM provides the PS, a type of end to end tunnel protection technology, to recover the interrupted services.

   The PS uses one tunnel to protect another tunnel. There is no relationship among the attributes of each tunnel in the protect group. For example, the protection tunnel with 10 Mbit/s bandwidth can protect a master tunnel with a requirement for 100 Mbit/s bandwidth.

MPLS OAM Detection Function

The basic process for MPLS OAM to detect the connectivity of a single LSP is as follows:

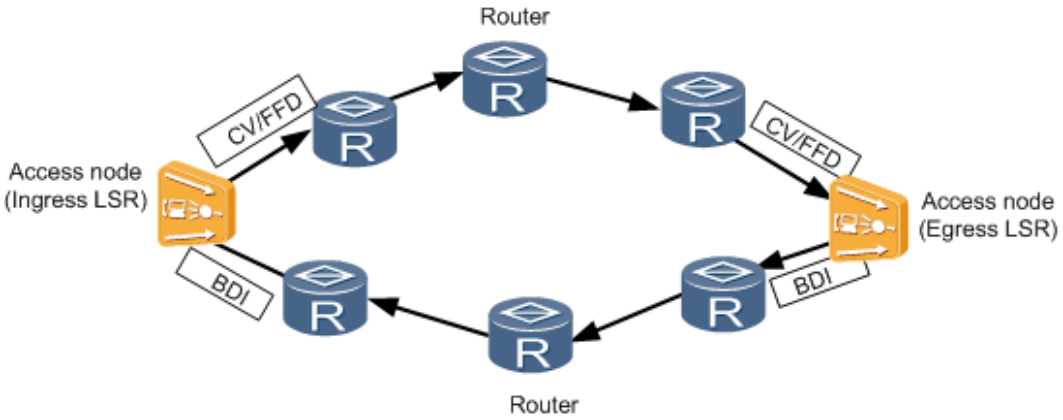
- The source transmits the CV/FFD packets to the destination through the detected LSP.
- The destination checks the correctness of the type and frequency information carried in the received detection packets and measures the number of correct and incorrect packets that are received within the detection period to monitor the connectivity of the LSP in real time.
- When the LSP fails, the destination detects the defect quickly and analyzes the defect type.

Bind a backward LSP to the detected LSP when configuring the OAM function for the detected LSP. A backward path is an LSP that has the opposite source and destination of the detected LSP, or a non-MPLS path that can be connected to the source and destination of the detected LSP.

After the destination detects a defect, the destination transmits the BDI packets that carry the defect information to the source through the backward path. The source learns about the status of the defect, and triggers the corresponding protection switching when the protect group is correctly configured. **Figure 6-6** shows the MPLS OAM CD.

 Copyright © Huawei Technologies Co., Ltd.

Figure 6-6 MPLS OAM CD



Working Modes of the MPLS OAM Protection Switching

The MPLS OAM protection switching aims at the entire LSP instead of one section or one node on the LSP. The route and bandwidth of the standby LSP for a specified active LSP are reserved.

Therefore, the protection switching is a thorough-assignment protection mechanism. To ensure that protection switching can be implemented effectively in all the possible cases that the active LSP fails, the standby LSP needs to use a physical path totally different from that of the active LSP.

The working mode of MPLS OAM protection switching is 1:1 protection mode. In this mode, each active LSP has a standby LSP.

- In normal conditions, data is transmitted through the active LSP and no traffic is transmitted through the standby LSP.
- When the destination detects a failure on the active LSP through the detection mechanism, the destination switches to the standby LSP, and then transmits the BDI packet to the source through the backward path, instructing the ingress to switch the traffic on the active LSP to the standby LSP. Thus, 1:1 protection switching is implemented.

6.6 MPLS TE Reliability

MPLS TE tunnels that transmit mission-critical services require high reliability. Access node supports the following network-level reliability.

- RSVP-TE FRR
- TE tunnel protection group
- CR-LSP backup

6.6.1 RSVP-TE FRR

RSVP-TE FRR is also called MPLS fast reroute. RSVP literally means the resource reservation protocol, TE means traffic engineering, and FRR means fast reroute.

 Copyright © Huawei Technologies Co., Ltd.

6.6.1.1 Introduction

Definition

The RSVP-TE FRR technology is applied to the MPLS TE network for implementing partial network protection. Specifically, when a certain link or node in the network fails, the LSP configured with FRR can automatically switch the data to the protection link.

To ensure the reliability of the MPLS network, the MPLS FRR technology is combined with the MPLS TE technology to provide LSPs with fast switching. In the MPLS FRR, a local backup path is created beforehand to protect the LSP from the impact of the link or node failure. When a failure occurs, the device that detects the failure can quickly switch the service from the faulty link to the backup path, thus reducing data loss.

Purpose

Quick response and prompt switching are the features of MPLS FRR. Such features ensure the smooth switching of service data and prevent service interruption. In addition, the head node of the LSP will look for a new path for establishing a new LSP and will switch the service to the new LSP. Before the new LSP is set up, the service data is forwarded through the protection path.

6.6.1.2 Principle

MPLS TE FRR

The basic principle of MPLS TE FRR is to protect one or more LSPs by using an LSP that is created beforehand. The LSP that is created beforehand is called the FRR LSP (bypass LSP), and the LSP that is protected is called a primary LSP. The purpose of MPLS TE FRR is to bypass the faulty link or node through the bypass LSP to protect the primary LSP. Creating the bypass LSP and primary LSP requires the participation of all the components of the MPLS TE system.

MPLS TE FRR is implemented based on RSVP TE and complies with RFC4090.

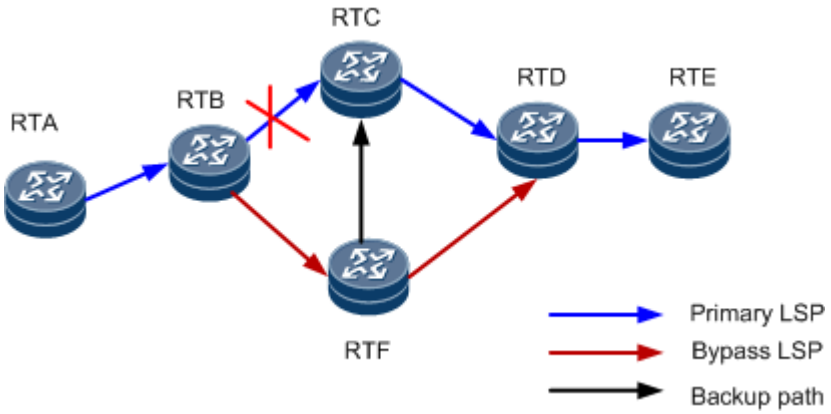
MPLS TE FRR can be implemented in the following two modes:

- Detour mode: This mode is also called the one-to-one backup mode. In this mode, one protection path is created to provide protection for each LSP. This protection path is called the detour LSP.
- Bypass mode: This mode is also called the facility backup mode. In this mode, one protection path provides protection for multiple LSPs. This protection path is called the bypass LSP.

The detour mode provides protection for each LSP, thus requiring more overheads. In the actual application, the bypass mode is more widely used. The MA5800 adopts the bypass mode. The following content of this topic mainly deals with the bypass mode. **Figure 6-7** illustrates the FRR function implemented in the bypass mode.

 Copyright © Huawei Technologies Co., Ltd.

Figure 6-7 FRR in the bypass mode

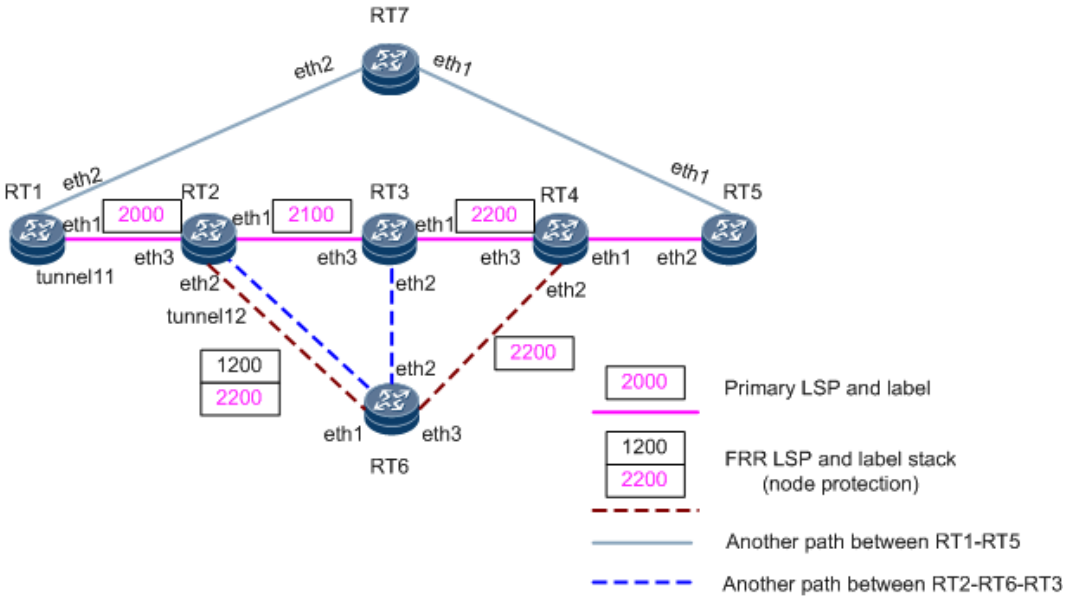


In **Figure 6-7**, the blue path is the primary LSP and the red path is the bypass LSP. When the link between RTB and RTC fails or when RTC fails, the data on the primary LSP is switched to the bypass LSP. The top layer of the packet header sent from RTB adopts the label assigned to RTB by RTF, and the egress label of RTC is also added to the label stack as the lower layer. The packet on the RTB-RTF-RTD LSP carries two labels. After receiving the packet, RTD finds the label assigned to RTF by RTD, and continues to use the label assigned to RTC by RTD for forwarding the packet.

Implementation Process of FRR in the Bypass Mode

Figure 6-8 illustrates the implementation process of FRR in the bypass mode.

Figure 6-8 FRR in the bypass mode



- Creating the primary LSP

 Copyright © Huawei Technologies Co., Ltd.

The primary LSP is created in the same way as an ordinary LSP is. The head node (RT1) sends the RSVP PATH message to downstream nodes one by one (RT1-RT2-RT3-RT4-RT5), and the end node (RT5) sends the RESV message to upstream nodes one by one. When processing the RESV message, each node assigns the label and reserves the resources for creating the LSP.

In the protocol draft, some flag bits in the SESSION_ATTRIBUT and RECORD_ROUTE objects are extended for FRR. The difference between the creating processes of the protected LSP and ordinary LSP lies in the processing of these flag bits. The flag bits added to the SESSION_ATTRIBUT object in the PATH message indicate whether the LSP needs partial protection, whether the label is recorded, and whether the bandwidth is protected.

The flag bits added to the RECORD_ROUTE object in the RESV message indicate whether the LSP is protected, whether the switching is enabled, whether the bandwidth is protected, and whether the node is protected.

The creating of the primary LSP is triggered through the manual configuration of a tunnel on the head node (RT1). Before the primary LSP is created, if the FRR attribute is specified for the LSP by a command, the partial protection flag will be added to and the label flag and the SE style flag will be recorded in the SESSION_ATTRIBUTE object in the RSVP PATH message. If bandwidth is also specified for the LSP, the RSVP will also add the bandwidth protection flag. After receiving the PATH message, through the local protection flag, the downstream node can determine that the LSP requires the FRR protection.

For the LSP that requires the FRR protection (determined according to the flag in the PATH message received), each node records the egress, LSR ID, and label of the RESV message in the RRO when sending the RESV message to the upstream node. Such information is passed on to each upstream node.

When receiving the RESV message for the first time, according to the information recorded in the RRO, each node selects a proper bypass LSP for the LSP to be protected (primary LSP). The process of selecting a proper bypass LSP for the primary LSP is called binding.

After the node performs the FRR binding calculation on the primary LSP, the node indicates whether the primary LSP has been protected in the RECORD_ROUTE object in the RESV message sent to the upstream node. If the primary LSP has been protected, the egress (eth1 of RT2) of the protected LSP and the egress (eth3 of RT2) of the RESV message are recorded. If the primary LSP is not protected, the corresponding flag bit in the RRO is reset, and only the egress (eth3 of RT2) of the RESV message is recorded. Binding calculation is not performed on the egress. All the flag bits in the RRO sent from the egress to the upstream node are reset.

The primary LSP requiring the FRR protection is created in a similar way to an ordinary LSP. The differences are that, in the creating process of the primary LSP, the binding calculation is added, and related flag bits and sub-objects are added to the PATH and RESV messages.

- Creating the bypass LSP

  A bypass LSP can be created in two modes: the manual mode and the automatic mode.

  In the manual mode, after a tunnel without the FRR attribute is specified for protecting a physical interface, the LSP corresponding to this tunnel becomes a bypass LSP. A manual bypass LSP (tunnel12 on RT2) is configured manually on the PLR (RT2). The configuration of a manual bypass LSP is similar to an

 Copyright © Huawei Technologies Co., Ltd.

ordinary LSP. The difference is that the bypass LSP cannot be configured with the FRR attribute. In other words, a bypass LSP cannot be a primary LSP at the same time. An LSP cannot be protected by itself.

The automatic mode of the bypass LSP simplifies the configuration of the manual mode. In the automatic mode, when the primary LSP requires the FRR protection, the PLR can select or automatically create a bypass LSP for protecting this primary LSP. A bypass LSP can protect multiple primary LSPs in so far as it meets the requirements of these primary LSPs.

A bypass LSP can protect multiple physical interfaces, but it cannot protect its own egress.

FRR can implement link protection or node protection. In the configuration of the bypass LSP, the links or nodes to be protected should be planned, and whether the link protection mode or node protection mode is to be adopted should be determined. Generally, node protection is a superior mode because it can protect the protected nodes and the links between the PLR and the protected nodes. If conditions permit, the customer tends to require node protection. Huawei device provides flexible protection modes. When node protection fails, the protection mode supported by Huawei device can automatically shift to link protection. When node protection becomes valid again, node protection will be adopted.

The bandwidth of the bypass tunnel is generally used for protecting the primary LSP. All the resources of the bypass tunnel are used only after the switching occurs. Make sure that the configured bandwidth of the bypass LSP is greater than or equal to the sum of the bandwidth required by all the protected LSPs. Otherwise, after FRR takes effect, the bypass LSP will fail to provide the protection that meets the service quality requirements.

A bypass LSP is generally in the idle state and does not carry data. If the bypass tunnel is required to forward data as well as protecting the primary LSP, sufficient bandwidth should be configured.

- Binding calculation

  Binding can refer to specifying a bypass tunnel for protecting a physical interface. Then, the bypass tunnel can be said to be bound to the physical interface.

  Binding can also refer to selecting a proper bypass LSP for protecting a primary LSP. Then, the primary LSP can be said to be bound to the bypass LSP. The binding calculation is a process of binding a primary LSP to the bypass LSP. The result derived from the binding calculation is the necessary data to be forwarded in the switching, such as the interface of the bypass tunnel, the egress and NHLFE of the bypass LSP, and the label assigned by the MP. If the binding calculation is successful, the node sends the RESV message to inform the upstream node that the primary LSP has been protected.

  The binding calculation must be completed before the switching occurs. In the following conditions, binding calculation is triggered:

  – When a primary LSP is created
  – When the system periodically calculates the binding relations of all the LSPs whose egress is the protected physical interface

  The binding calculation always uses the known information of a primary LSP to traverse the bypass LSPs on the egress through which the primary LSP is protected, thus to find a most suitable bypass LSP. If automatic bypass LSP is

 Copyright © Huawei Technologies Co., Ltd.

supported, when a suitable bypass LSP is not found, the system will automatically try to create a bypass LSP for protecting the primary LSP.

When the primary LSP is created, the interface address of each node is recorded. The CSPF can obtain the corresponding LSR ID according to the interface address. Hence, the LSR ID of the next hop (NHP) or next next hop (NNHOP) of the primary LSP is known. When the primary LSP is created, the RRO records the LSR ID of each hop. If the egress LSR ID and the NHP LSR ID of a bypass LSP are the same, link protection can be realized; if the egress LSR ID and NNHOP LSR ID of a bypass LSP are the same, node protection can be realized.

If the bandwidth of a primary LSP is 0, it can be protected only by a bypass LSP whose bandwidth is 0. After a primary LSP comes into the protection of a bypass LSP whose bandwidth is 0, the protection count of this bypass LSP is plus 1. If the bandwidth of a primary LSP is not 0, it can be protected only by a bypass LSP with sufficient remaining bandwidth. The initial remaining bandwidth of a bypass LSP whose bandwidth is not 0 is the configured value. Each time a primary LSP comes into the protection of the bypass LSP, the remaining bandwidth of the bypass LSP is minus the bandwidth of the primary LSP.

When multiple bypass LSPs are available for protecting a primary LSP, the following priority is adopted:

- Node protection is prior to link protection.
- If the bandwidth of the primary LSP is 0, a bypass LSP whose bandwidth is 0 is selected. If the bandwidth of the primary LSP is not 0, the bypass LSP whose remaining bandwidth is greater than or equal to the bandwidth of the primary LSP is selected.

The result derived from the binding calculation contains the following items, which are used for sending the data and signaling message from the bypass tunnel after the switching.

- Protection type (link protection or node protection), and the LSR ID of the MP.
- The label assigned to the last hop by the MP. This label is the label corresponding to the MP LSR ID in the RRO of the primary LSP.
- Egress and NHLFE of the bypass tunnel.

The binding calculation result is saved and can be immediately used when partial failure occurs. This is why MPLS TE FRR can respond quickly to failure.

- Failure detection

  The purpose of failure detection is to detect the failure of a link (RT2-RT3) or a node (RT3) as soon as possible so that switching can be triggered to reduce packet loss. Failure detection does not specifically distinguish between a link and a node, and the result of failure detection is presented as "interface failure" (eth1 of RT2).

  The "interface failure" triggers the FRR switching on all the LSPs that use the interface as the egress. If an LSP has been determined by the binding calculation to be in the link protection, the LSP will switch to link protection. If the actual failure is a node failure, the switching fails. As a result, the LSP is deleted. If an LSP has been determined by the binding calculation to be in the node protection, the LSP will switch to node protection. If the actual failure is a link failure and even if the next hop is available, the next hop will be skipped by the bypass tunnel.

 Copyright © Huawei Technologies Co., Ltd.

Certain link or node failures can be detected by the link layer protocol. The detection speed of the link layer protocol is directly related to the interface type. Other link or node failures are detected through the hello mechanism of the RESV. The detection speed of the hello mechanism is relatively slow.

The hello function can be enabled on each physical interface that needs protection and on its interconnected interface. Then, the hello message and the response will be sent between the two routers periodically. In case of a link or node failure, the hello message or the response is lost. When the messages are lost for three successive times, it is regarded that a failure occurs.

- Switching

  Switching refers to adopting the bypass LSP for sending the data and RSVP messages that used to be sent through the primary LSP. When the interface (eth1 of RT2) is shut down by a command or when "interface failure" (of eth1 of RT2) is detected through the failure detection mechanism, switching is triggered. In the switching, the data and signaling of the protected LSPs on the faulty interface are switched to the bypass LSP for sending, and the upstream node is informed that switching occurs.

  Switching first occurs on the forwarding components. During the binding calculation by the forwarding component involved in the switching, the inner label (2200) required for the forwarding has been saved in the NHLFE. Now, it only needs to indicate that the LSP has been switched, and the data can be forwarded through the bypass tunnel.

  Then, the node will respond to the switching event through the RESV message. For the LSP that has been bound to the bypass LSP, the node sends the upstream node the RSVP PathError message with the switching flag bit. The bypass tunnel is mainly used for temporary protection. The head node will properly process the LSPs that have been switched. If an LSP is not bound to the bypass LSP, the node directly sends the RSVP ResvTear message to inform the upstream node to delete the LSP.

- Maintenance of the LSP after the switching

  After the switching, the original link is not available. To prevent the LSP from being deleted after timeout, the information between the PLR (RT2) and the MP (RT4) needs to be refreshed through RSVP messages.

  After being modified, the PATH message is sent to the MP through the bypass tunnel (Tunnel12 of RT2). After receiving the PATH message, the MP confirms itself as an MP. Then, the RESV message is modified and forwarded to the PLR through the IP addresses of multicast hops (RT4-RT6-RT2).

  After the switching, the message sent from the PLR to the upstream node is also changed. That is, the address of the egress (eth2 of RT2) of the bypass LSP is added to the RRO.

  After the switching, the sending path of the PTEAR, RERR, RTEAR and PERR messages of the primary LSP are changed accordingly.

  After the switching in node protection, the protected node (RT3) may send the PATHTEAR message to the downstream node because the PATH message times out. In this case, the MP (RT4) ignores this message. In addition, in the switching, the MP sends the ResvTear message from the ingress (eth3 of RT4) of the original LSP. Thus, the protected node (RT3) will release the corresponding resource as soon as possible.

 Copyright © Huawei Technologies Co., Ltd.

- Re-optimization

  Re-optimization refers to calculating the path for a created LSP at the preset intervals. According to the calculated path, the router initiates the creating of a new LSP. After the new LSP is created, the original LSP is deleted, and the data of the original LSP tunnel is switched to the new LSP for forwarding.

  Re-optimization can be configured for each LSP tunnel. After the LSP is created, re-optimization is enabled.

  In the case of FRR, another function of re-optimization is to restore the tunnel (Tunnel1 of RT1) protected by the bypass LSP to the normal state. This is because the FRR protection is temporary. Therefore, a tunnel with the FRR attribute is generally configured with re-optimization. When the primary LSP has not switched, a new LSP is created only when the path calculated through re-optimization is different from the original path. When the primary LSP has switched, a new LSP is created even when the path calculated through re-optimization is the same as the original path.

  A bypass LSP that has been bound to a physical interface can also be re-optimized. The bypass LSP, however, cannot be re-optimized if a primary LSP already switches to this bypass LSP. After a bypass LSP is re-optimized, the binding relations between the bypass LSP and the primary LSPs are refreshed.

  Before the primary LSP is switched, the data forwarding is the same as that of an ordinary LSP; after the primary LSP is switched to the bypass tunnel, the data is forwarded through the bypass tunnel to the MP.

  When the primary LSP is successfully bound to the bypass LSP, the NHLFE entry and the inner label (2200, the label assigned to the last upstream node by the MP) of the bypass LSP are recorded in the NHLFE entry of the primary LSP. In the switching, the forwarding component sets the switching flag bit in the NHLFE entry of the primary LSP.

  When the packet arrives at the PLR, the forwarding component searches for the NHLFE entry to the primary LSP. If switching has not occurred, the component performs label switching and data forwarding; if the switching flag bit is found in the NHLFE entry, the component continues searching for the NHLFE entry to the corresponding bypass LSP. After finding the NHLFE entry, the component adds inner label 2200 to the label stack, and performs forwarding according to the information of the NHLFE entry of the bypass LSP.

  At the egress of the bypass tunnel (or at the last but one hop), inner label 2200 is removed from the label stack, and then MP can perform forwarding by using the original label 2200. The inner label may be used on different interfaces of the MP. Therefore, the MP must assign a label to each platform.

  As previously mentioned, certain failures are detected at the link layer. After a failure is detected at the link layer, the forwarding component can reset the switching flag bit in the NHLFE entry of the primary LSP if the failure recovers before a corresponding failure occurs at the upper layer. Hence, the data of the primary LSP is still forwarded through the original path, and the switching flag in the RESV message is not processed.

  One thing should be noted that, after the switching, the RSVP message from the PLR to the MP is sent through the bypass tunnel. In other words, the message is forwarded as a common IP packet through the MPLS tunnel. The RSVP message from the MP to the PLR is forwarded as a common IP packet.

 Copyright © Huawei Technologies Co., Ltd.

6.6.2 TE Tunnel Protection Group

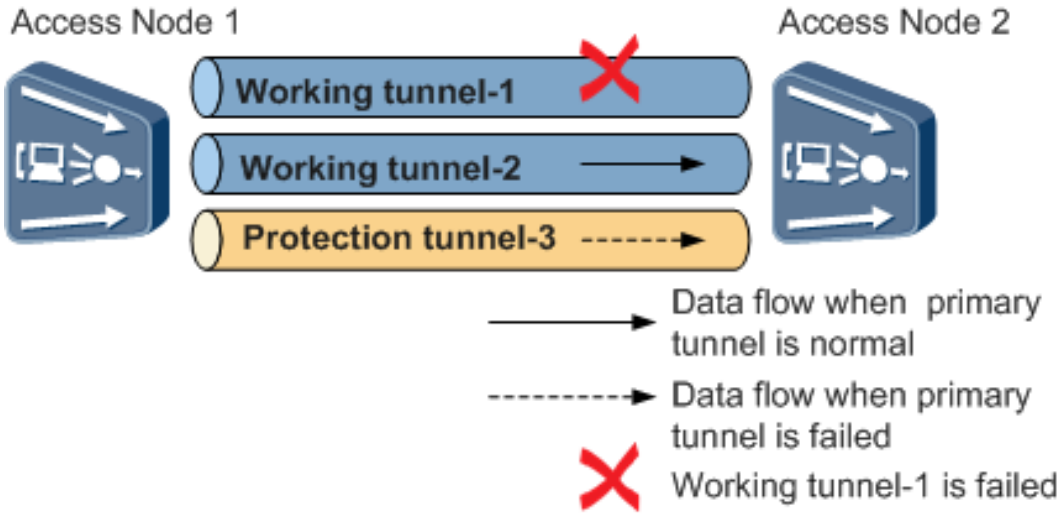
A tunnel protection group protects end-to-end MPLS TE tunnels. If a working tunnel in a protection group fails, traffic switches to a protection tunnel, minimizing traffic interruptions.

6.6.2.1 Related Concepts

As shown in the **Figure 6-9**, concepts related to a tunnel protection group are as follows:

- Working tunnel: a tunnel to be protected.
- Protection tunnel: a tunnel that protects a working tunnel.
- Protection switchover: switches traffic from a faulty working tunnel to a protection tunnel in a tunnel protection group, which improves network reliability.

Figure 6-9 Tunnel protection group



Primary tunnel-1 and tunnel-2, and the bypass tunnel-3 are established on the ingress Access Node shown in the **Figure 6-9**.

Tunnel-3 is specified as a protection tunnel for primary tunnels tunnel-1 and tunnel-2 on Access Node. If the configured fault detection mechanism on the ingress detects a fault in tunnel-1, traffic switches to tunnel-3. Access Node attempts to reestablish tunnel-1. If tunnel-1 is successfully established, traffic switches back to the primary tunnel.

6.6.2.2 Principle

Implementation

A TE tunnel protection group uses a configured protection tunnel to protect traffic on the working tunnel to improve tunnel reliability. To ensure the improved

 Copyright © Huawei Technologies Co., Ltd.

performance of the protection tunnel, the protection tunnel must exclude links and nodes through which the working tunnel passes during network planning.

Table 6-2 shows the implementation procedure of a tunnel protection group.

Table 6-2 Implementation procedure of a tunnel protection group

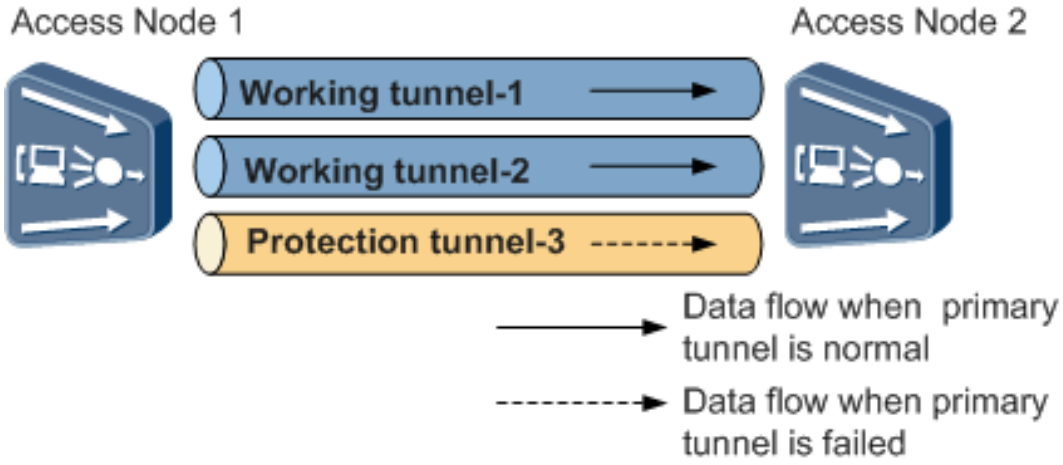
Sequence Number	Process	Description
1	Establishment	The working and protection tunnels must have the same ingress and destination address. The protection tunnel is established in the same procedure as a regular tunnel. The protection tunnel can use attributes that differ from those for the working tunnel. Ensure that the working and protection tunnels are established over different paths as much as possible. **NOTE** • A protection tunnel cannot be protected or enabled with TE FRR. • Attributes for a protection tunnel can be configured independently of those for the working tunnel, which facilitates the network planning.
2	Binding between the working and protection tunnels	The protection tunnel is bound to the tunnel ID of the working tunnel so that the two tunnels form a tunnel protection group.
3	Fault detection	In addition to MPLS TE's own detection mechanism, MPLS OAM and BFD for CR-LSP are used to detect faults in a tunnel protection group to speed up protection switching.
4	Protection switching	The tunnel protection group supports either of the following protection switching modes: • Manual switching: Traffic is forcibly switched to the protection tunnel. • Automatic switching: Traffic automatically switches to the protection tunnel if the working tunnel fails. A time interval can be set for automatic switching.
5	Switchback	After a traffic switchover is implemented, the ingress attempts to reestablish the working tunnel. If the working tunnel is reestablished, the ingress can switch traffic back to the working tunnel or still forward traffic over the protection tunnel.

 Copyright © Huawei Technologies Co., Ltd.

Protection mode

A tunnel protection group works in either 1:1 or N:1 mode. The 1:1 mode enables a protection tunnel to protect only a single working tunnel. The N:1 mode enables a protection tunnel to protect more than one working tunnel. The following figure shows the N:1 protection mode.

Figure 6-10 N:1 protection mode



Differences Between CR-LSP Backup and a Tunnel Protection Group

CR-LSP backup and a tunnel protection group are both E2E protection mechanisms for MPLS TE. **Table 6-3** shows the comparison between these two mechanisms.

Table 6-3 Comparison between CR-LSP backup and a tunnel protection group

Item	CR-LSP Backup	Tunnel Protection Group
Object to be protected	Primary and backup CR-LSPs are established on the same tunnel interface. A backup CR-LSP protects traffic on a primary CR-LSP.	One tunnel protects traffic over another tunnel in a tunnel protection group.
TE FRR	A primary CR-LSP supports TE FRR. A backup CR-LSP does not support TE FRR.	A working tunnel supports TE FRR. A protection tunnel does not support TE FRR.

 Copyright © Huawei Technologies Co., Ltd.

Item	CR-LSP Backup	Tunnel Protection Group
LSP attributes	Primary and backup CR-LSPs have the same attributes, except for the TE FRR attribute.	The attributes of one tunnel in a tunnel protection group are independent of the attributes of the other tunnel. For example, a protection tunnel with no bandwidth can protect traffic on a working tunnel that has a bandwidth.
Protection mode	A 1:1 protection mode is supported. Each primary CR-LSP is protected by a backup CR-LSP.	An N:1 protection mode is supported. Many tunnels share one protection tunnel. If any protected tunnel fails, traffic switches to the protection tunnel.

6.6.3 CR-LSP Backup

CR-LSP backup techniques protect E2E MPLS TE tunnels. If the ingress detects that the primary CR-LSP is unavailable, the ingress switches traffic to a backup CR-LSP. After the primary CR-LSP recovers, traffic switches back.

6.6.3.1 Related Concepts

CR-LSP backup functions include hot standby, ordinary backup, and the best-effort path function. CR-LSP backup functions are as follows:

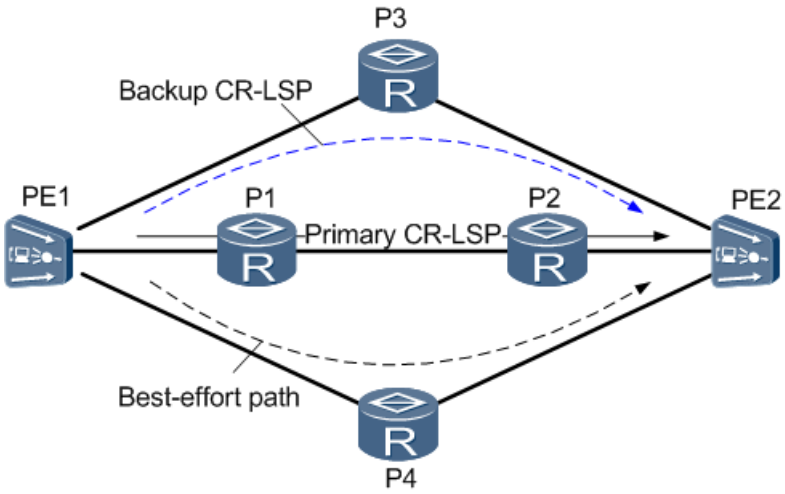
- Hot standby: A hot-standby CR-LSP is established immediately after a primary CR-LSP is created. If the primary CR-LSP fails, the hot-standby CR-LSP takes over traffic from the primary CR-LSP. After the primary CR-LSP recovers, traffic switches back.
- Ordinary backup: An ordinary backup CR-LSP can be established only after a primary CR-LSP fails. The ordinary backup CR-LSP takes over traffic if the primary CR-LSP fails. After the primary CR-LSP recovers, traffic switches back.
- Best-effort path

  If both the primary and backup CR-LSPs fail, a best-effort path is established and takes over traffic.

  For example, the primary CR-LSP is established over the path PE1 → P1 → P2 → PE2, and the backup CR-LSP is established over the path PE1 → P3 → PE2 shown in **Figure 6-11**. If both CR-LSPs fail, PE1 establishes a best-effort path PE1 → P4 → PE2 to take over traffic.

 Copyright © Huawei Technologies Co., Ltd.

Figure 6-11 Best-effort path



NOTE

A best-effort path has no bandwidth reserved for traffic, but has an affinity and a hop limit configured as needed.

6.6.3.2 Principle

Implementation

The procedure of CR-LSP backup is as follows:

1. CR-LSP backup is deployed.

   Plan the paths, bandwidth values, and deployment modes. **Table 6-4** lists CR-LSP backup deployment items.

 Copyright © Huawei Technologies Co., Ltd.

Table 6-4 CR-LSP backup deployment

Item	Hot Standby	Ordinary Backup	Best-Effort Path
Path	Determine whether the primary and hot-standby CR-LSPs entirely or partially overlap. A hot-standby CR-LSP can be established over an explicit path. A hot-standby CR-LSP supports the following attributes: ● Explicit path ● Affinity ● Hop limit	Allowed to use the path of the primary CR-LSP in all scenarios. An ordinary backup CR-LSP supports the following attributes: ● Explicit path ● Affinity ● Hop limit	Automatically calculated by the ingress. A best-effort path supports the following attributes: ● Affinity ● Hop limit
Bandwidth	A hot-standby CR-LSP and a primary CR-LSP have the same bandwidth by default.	An ordinary backup CR-LSP and a primary CR-LSP have the same bandwidth.	A best-effort path is only a protection path that does not have reserved bandwidth.
Configuration combination	A hot-standby CR-LSP can be used together with a best-effort path.	An ordinary CR-LSP can only be used alone.	–

2. Fault detection is implemented.

   CR-LSP backup supports the RSVP-TE fault advertisement mechanism, who sends signaling packets to detect faults at a low speed.

3. A traffic switchover is implemented.

   If a primary CR-LSP fails, the ingress attempts to switch traffic from the primary CR-LSP to a hot-standby CR-LSP. If the hot-standby CR-LSP is unavailable, the ingress attempts to switch traffic to an ordinary backup CR-LSP. If the ordinary backup CR-LSP is unavailable, the ingress attempts to switch traffic to a best-effort path.

4. A traffic switchback is implemented.

 Copyright © Huawei Technologies Co., Ltd.

Traffic switches back to a path based on the available CR-LSPs. Traffic will switch first to the primary CR-LSP, which has the highest priority. If the primary CR-LSP is unavailable, traffic will switch to the hot-standby CR-LSP. The ordinary CR-LSP has the lowest priority.

Overlapping Path for a Hot-standby CR-LSP

The overlapping path function can be configured for a hot-standby CR-LSP. The path of the hot-standby CR-LSP can overlap the path of a primary CR-LSP in all scenarios.

Coexistence of CR-LSP Backup and TE FRR

1. CR-LSP backup functions can be used together with TE FRR.
   - Hot standby and TE FRR: If TE FRR detects a link fault, traffic switches to a TE FRR bypass CR-LSP and then to a hot-standby CR-LSP.
   - Ordinary backup and TE FRR: If TE FRR detects a link fault, traffic switches to a TE FRR bypass CR-LSP. If both the primary and TE FRR bypass CR-LSPs fail, an ordinary backup CR-LSP is established and takes over traffic.
2. CR-LSP backup can be associated with TE FRR.

   The association improves tunnel security. The association provides the following functions based on backup modes:

   - Association between an ordinary backup CR-LSP and a TE FRR bypass CR-LSP provides the following functions:

     If a protected link or node fails, traffic switches to a bypass CR-LSP. The ingress attempts to reestablish the primary CR-LSP, while attempting to establish an ordinary backup CR-LSP.

     If the ordinary backup CR-LSP is established successfully before the primary CR-LSP is restored, traffic switches to the ordinary backup CR-LSP.

     After the primary CR-LSP recovers, traffic switches back to the primary CR-LSP.

     If the ordinary backup CR-LSP fails to be established, and the primary CR-LSP does not recover, the traffic still passes through the bypass CR-LSP.
   - Association between a hot-standby CR-LSP and a TE FRR bypass CR-LSP provides the following functions:

     If a hot-standby CR-LSP is Up and a protected link or node fails, traffic switches to a TE FRR bypass CR-LSP and then immediately switches to the hot-standby CR-LSP. At the same time, the ingress attempts to restore the primary CR-LSP.

     If the hot-standby CR-LSP is Down, the traffic switching procedure is the same as that when the ordinary backup is used.

   Association between ordinary backup CR-LSPs and TE FRR is recommended. An ordinary backup CR-LSP without additional bandwidth needed is established only after the primary CR-LSP enters the FRR-in-use state. Although the primary CR-LSP is Up, the system attempts to establish a hot-standby CR-LSP with additional bandwidth needed.

 Copyright © Huawei Technologies Co., Ltd.

6.7 Configuring the MPLS Service

This topic describes the MPLS technology and how to configure the MPLS service on the MA5800.

Basic concept

- The path that an FEC traverses in an MPLS network is called LSP. The LSP, whose function is the same as the virtual circuit in ATM and frame relay, is a unidirectional path from the ingress to the egress. Each node on the LSP is an LSR.
- The static LSP is the label forwarding path manually set up for label distribution to each FEC.
- The dynamic LSP is the label forwarding path dynamically established through the label distribution protocol (LDP or RSVP-TE).

Configuration logic

In the MPLS configuration, the core is to configure the LSP and the second is to configure fault detection and protection for the LSP. At the same time, According to the protocol for creating LSPs, LSPs are categorized as static LSP, LDP LSP, and RSVP-TE LSP.

Therefore, configure MPLS as follows:

1. Configure LSPs.
   - Configure a static LSP.
   - Configure an LDP LSP.
   - Configure an RSVP-TE LSP.
2. Configure LSP protection. Configure the MPLS OAM.

6.7.1 Configuring the Static LSP

Static LSP is configured manually. A static LSP can work in the normal state only when all the LSRs along the static LSP are configured.

Prerequisites

1. The IP address of the loopback interface must be configured.
2. The LSR ID must be configured.
3. The global MPLS, VLAN MPLS, and VLAN interface MPLS must be enabled.
4. A static or dynamic route must be successfully configured on each device in the network (so that LSRs can reach each other through the IP route).

Context

The administrator needs to manually distribute labels to each LSR when configuring the static lsp. Principle: The out label value of a node must be equal to the in label value of its next node. LSRs on a static LSP cannot perceive the entire LSP. Therefore, static LSP is a local concept.

 Copyright © Huawei Technologies Co., Ltd.

The MA5800 can function as a label switching edge router (LER) or a label switching router (LSR). According to the position of the LER or LSR in a network, the configuration of the static LSP involves the ingress configuration, transit node configuration, and egress configuration.

An LSP corresponds to a unidirectional forwarding path. To ensure bidirectional communication of the MPLS service, two static LSPs are required. The two LSPs have opposite directions. Their ingress and egress are reverse. Their transit nodes can be the same or different according to the networking requirements, or even free of being configured.

Procedure

- When the MA5800 functions as an LER, configure the static LSP as follows:
  a. Run the **static-cr-lsp ingress** command to configure the ingress parameters of a static LSP.

     An LER is generally located at the edge of an MPLS network. The PE or PTN device can be considered an LER.

     Format:

     static-cr-lsp ingress { **tunnel-interface tunnel** *interface-number* | **tunnel-name** } **destination** *destination-address* { **nexthop** *next-hop-address* | **outgoing-interface** *interface-type interface-number* } * **out-label** *out-label* [**bandwidth ct0** *bandwidth*]

     - You can create a static LSP by using the LSP name or the tunnel. To create a static LSP by using the tunnel, you must run the **interface tunnel** command to create a tunnel interface and then configure its attributes.
     - **destination** *ip-addr*: Indicates the destination IP address of the LSP, that is, the loopback interface IP address of the PE or PTN device.
     - **nexthop** *ip-addr*: Indicates the next hop IP address, that is, the VLAN interface IP address of the adjacent LSR.
     - **out-label** *out-label*: Indicates the out label value, which must be the same as the in label value of the downstream LSR.

  b. Run the **static-cr-lsp egress** command to configure the egress parameters of a static LSP.

     Format:

     static-lsp egress *lsp-name* **incoming-interface vlanif** *vlanid* **in-label** *in-label*[**lsrid** *ingress-lsr-id* **tunnel-id** *tunnel-id*]

     - In the egress configuration of a static LSP, only a VLAN interface can be used as the ingress interface.
     - **in-label** *in-label*: Indicates the in label value of the egress, which must be the same as the out label value of the upstream LSR.

  c. Run the **display mpls static-cr-lsp** command to query the configuration of a static LSP.

 Copyright © Huawei Technologies Co., Ltd.

- When the MA5800 functions as an LSR, configure the static LSP as follows:
  a. Run the **static-cr-lsp transit** command to configure the transit node parameters of a static LSP.

     An LSR is generally located in the middle of an MPLS network. The P device can be considered an LSR that forwards MPLS labels.

     Format:

     static-cr-lsp transit *lsp-name* **incoming-interface** *interface-type interface-number* **in-label** *in-label* { **nexthop** *next-hop-address* | **outgoing-interface** *interface-type interface-number* } * **out-label** *out-label* [**ingress-lsrid** *ingress-lsrid* **egress-lsrid** *egress-lsrid* **tunnel-id** *tunnel-id*] [**bandwidth** [**ct0**] *bandwidth*] [**description** *description-text*]

     - The ingress interface of the transit node on a static LSP can only be the VLAN interface, that is, the VLAN interface of the upstream egress.
     - **in-label** *in-label*: Indicates the in label value of the transit node, which must be the same as the out label value of the upstream ingress.
     - **nexthop** *next-hop-address*: Indicates the next hop IP address, that is, the VLAN interface IP address of the adjacent LSR.
     - **out-label** *out-label*: Indicates the out label value of the transit node, which must be the same as the in label value of the downstream LSR.

     NOTICE

     Because the LSP is unidirectional, you must configure the transit node parameters twice with opposite directions to ensure bidirectional communication of the MPLS service.

  b. Run the **display mpls static-cr-lsp** command to query the configuration of a static LSP.

----End

Example

When the MA5800 functions as an LER, to configure the ingress and egress of a static LSP, set the parameters as follows:

- Ingress node name of the static LSP: lsp1; egress name of the static LSP: lsp2
- IP address of local VLAN interface 100: 10.1.1.2/24
- Destination IP address of the LSP: 10.3.3.3/32
- Out label: 8200; in label: 8300
- Next hop IP address: 10.1.1.3

```
huawei(config)#static-cr-lsp ingress lsp1 destination 10.3.3.3 nexthop 10.1.1.3 out-label 8200
huawei(config)#static-cr-lsp egress lsp2 incoming-interface vlanif 100 in-label 8300
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#display mpls static-cr-lsp
{ <cr>|STRING<1-19>|exclude<K>|include<K>|verbose<K>||<K> }:
  Command:

        display mpls static-lsp
TOTAL         :       2       STATIC LSP(S)
UP            :       0       STATIC LSP(S)
DOWN          :       2       STATIC LSP(S)
Name              FEC              I/O Label    I/O If              Status
lsp1              10.3.3.3           NULL/8200    -/-                 Down
lsp2              -                  8300/NULL    vlanif100/-         Down
```

When the MA5800 functions as an LSR, to configure the transit node parameters of a static LSP, set the parameters as follows:

- LSP name of the transit node in the positive direction: lsp1; LSP name of the transit node in the negative direction: lsp2
- IP address of local VLAN interface 100: 10.1.1.2/24
- IP address of local VLAN interface 200: 10.2.2.2/24
- Out label in the positive direction: 8200; in label in the positive direction: 8300
- Out label in the negative direction: 8200; in label in the negative direction: 8300
- Next hop IP address in the positive direction: 10.2.2.3
- Next hop IP address in the negative direction: 10.1.1.3

```
huawei(config)#static-cr-lsp transit lsp1 incoming-interface vlanif 100 in-label 82 00 nexthop 10.2.2.3
out-label 8300
huawei(config)#static-lsp transit lsp2 incoming-interface vlanif 200 in-label 83 00 nexthop 10.1.1.2
out-label 8200
huawei(config)#display mpls static-cr-lsp
{ <cr>|STRING<1-19>|exclude<K>|include<K>|verbose<K>||<K> }:
  Command:

        display mpls static-lsp
TOTAL         :       2       STATIC LSP(S)
UP            :       0       STATIC LSP(S)
DOWN          :       2       STATIC LSP(S)
Name              FEC              I/O Label    I/O If              Status
lsp1              -/-              8200/8300    vlanif100/-         Down
lsp2              -/-              8300/8200    vlanif200/-         Down
```

6.7.2 Configuring the LDP LSP

Set up an MPLS LDP session between LSRs along the LSP. After the MPLS LDP session is set up, the LDP LSP is automatically created.

Prerequisites

1. The IP address of the loopback interface must be configured.
2. The LSR ID must be configured.
3. The VLAN for MPLS label forwarding must be created.
4. Global MPLS must be enabled.
5. A static or dynamic route must be successfully configured on each device in the network (so that LSRs can reach each other through the IP route).

Context

- The MA5800 supports LDP and RSVP-TE, both of which generate dynamic LSPs.

 Copyright © Huawei Technologies Co., Ltd.

- LDP is a standard MPLS label distribution protocol defined by IETF. LDP, which is mainly used to distribute labels for the negotiation between LSRs to set up label switching paths (LSPs), regulates various types of information for the label distribution process, and the related processing. The LSRs form an LSP that crosses the entire MPLS domain according to the local forwarding table, which correlates the in label, network hop node, and out label of each specific FEC.

Procedure

Step 1 Configure the MPLS LDP session.

The MPLS-LDP session is used for information exchange such as label mapping and release between LSRs. The MPLS-LDP session is classified into two types:

- Local LDP session: Two LSRs between which a session is set up are connected directly.
- Remote LDP session: Two LSRs between which a session is set up are not connected directly. Remote LDP sessions are mainly set up between nonadjacent LSRs. They can also be set up between adjacent LSRs.

NOTE

If local adjacency with the specified remote peer exists, remote adjacency cannot be set up; if remote adjacency exists and local adjacency is set up for the remote peer, the remote peer will be deleted. In other words, only one session can exist between two LSRs and a local LDP session takes priority over a remote LDP session.

- Configure the local LDP session.
  a. In the global config mode, run the **mpls ldp** command to enable global MPLS LDP.
  b. Run the **interface vlanif** command to enter the VLAN interface mode.
  c. In the VLAN interface mode, run the **mpls** command to enable the MPLS function of the VLAN interface and run the **mpls ldp** command to enable the MPLS LDP function of the VLAN interface.

     NOTE

     The VLAN 1 is the system default VLAN. All the upstream ports have been added to this VLAN by default. Do not use this VLAN as the MPLS VLAN or enable the MPLS function on this VLAN.

  d. Run the **quit** command to quit the VLAN interface mode.
- Configure the remote LDP session.
  a. In the global config mode, run the **mpls ldp** command to enable global MPLS LDP.
  b. Run the **mpls ldp remote-peer** command to create an LDP remote peer and then enter the remote peer mode.
  c. Run the **remote-ip** command to configure the IP address of the LDP remote peer.

     NOTE

     The IP address of the remote LDP peer should be the LSR ID of the remote LSR. When the LSR ID is used as the transmission address of a remote peer, two remote peers set up a TCP connection between them using the LSR ID as the transmission address.

 Copyright © Huawei Technologies Co., Ltd.

d. (Optional) Run the **mpls ldp advertisement** command to set the label distribution mode to DoD (downstream on demand) or DU (downstream unsolicited, default).

In a network with a large scale, it is recommended to set the mode to DoD to reduce unnecessary MPLS forwarding entries.

Step 2 (Optional) Configure the LDP MTU signaling function.

Run the **mtu-signalling** command to enable the sending of the MTU type, length, and value (TLV). This enables the LDP to automatically calculate and negotiate the minimum MTU value for all ports on each LSP. In this way, the MPLS determines the size of the MPLS forwarding packet at the ingress according to the minimum MTU, thereby avoiding the forwarding failure on transit nodes caused by oversize packets at the ingress.

By default, the LDP MTU signaling is enabled.

Step 3 (Optional) Configure the route trigger policy for setting up an LSP.

Run the **lsp-trigger host** command to configure the route trigger policy for setting up an LSP. The default route trigger policy is used to set up an LSP by triggering the LDP through the host address. To modify the default route trigger policy, run this command.

NOTE

It is recommended that you configure the route trigger policy for setting up an LSP to host (default), that is, the host route triggers the LDP to set up an LSP. In this way, the setup of useless LSPs can be prevented.

Step 4 (Optional) Configure the trigger policy set up by the transit LSP.

Run the **propagate mapping** command to filter certain routes received by the LDP by using the IP prefix table. Only the route that matches the specified IP prefix table is used by the local LDP for creating the transit LSP. By default, the LDP does not filter the received routes when creating the transit LSP.

Step 5 Query the relevant information about the LDP LSP configuration.

- Run the **display mpls ldp lsp** command to query the relevant information about the created LDP LSP.
- Run the **display mpls ldp session** command to check whether the created remote MPLS LDP session is in the normal (operational) state.
- Run the **display mpls interface** command to check whether the MPLS interface is in the normal (up) state.

----End

Example

To configure an LDP LSP between two adjacent LSRs by using VLAN interface 200 as the MPLS forwarding interface and using default values for other parameters, do as follows:

```
huawei(config)#mpls ldp
huawei(config-mpls-ldp)#quit
huawei(config)#interface vlanif 200
huawei(config-if-vlanif200)#mpls ldp
huawei(config-if-vlanif200)#quit
huawei(config)#display mpls interface vlanif 200
```

 Copyright © Huawei Technologies Co., Ltd.

```
{ <cr>|verbose<K> }:

  Command:
        display mpls interface vlanif 200
Interface            Status    TE Attr   LSP Count  CRLSP Count Effective MTU
vlanif200             Down      Dis       0          0           1500
```

To configure an LDP LSP between two nonadjacent LSRs by configuring the local lsr-id to 10.3.3.3, configuring the remote lsr-id to 10.5.5.5, and using default values for other parameters, do as follows:

```
huawei(config)#mpls ldp
huawei(config-mpls-ldp)#quit
huawei(config)#mpls ldp remote-peer session1
huawei(config-mpls-ldp-remote-session1)#remote-ip 10.5.5.5
huawei(config-mpls-ldp-remote-session1)#quit
huawei(config)#display mpls ldp remote-peer
{ <cr>|peer-id<K>|string<S><Length 1-32>||<K> }:
                                                                  Command:
       display mpls ldp remote-peer
                                                                                  LDP Remote Entity
Information
 ------------------------------------------------------------------------------
 Remote Peer Name  : session1
 Remote Peer IP    : 10.5.5.5            LDP ID        : 10.1.1.1:0
 Transport Address : 10.1.1.1            Entity Status : Active
                                                                  Configured Keepalive Hold Timer : 45
Sec
 Configured Keepalive Send Timer : ---
 Configured Hello Hold Timer     : 45 Sec
 Negotiated Hello Hold Timer     : 45 Sec
 Configured Hello Send Timer     : ---
 Configured Delay Timer          : 10 Sec
 Hello Packet sent/received      : 0/0
 Label Advertisement Mode        : Downstream Unsolicited
 Remote Peer Deletion Status     : No
 Auto-config                     : ---
 ------------------------------------------------------------------------------
 TOTAL: 1 Peer(s) Found.
```

6.7.3 Configure an RSVP-TE LSP

MPLS TE is a technology that integrates TE with MPLS. Through the MPLS TE technology, you can create an LSP tunnel to a specified path, to reserve resources and implement re-optimization.

Prerequisites

1. The IP address of the loopback interface must be configured.
2. The LSR ID must be configured.
3. The VLAN for MPLS label forwarding must be created.
4. Global MPLS and VLAN MPLS must be enabled.
5. The OSPF protocol must be successfully configured on each device in the network (the host route of each port must be successfully advertised).

Context

- To create constraint-based LSPs in MPLS TE, RSVP is extended. The extended RSVP signaling protocol is called the RSVP-TE signaling protocol.
- MPLS TE creates the LSP tunnel along a specified path through RSVP-TE and reserves resources. Thus, carriers can accurately control the path that traffic

traverses to avoid the node where congestion occurs. This solves the problem that certain paths are overloaded and other paths are idle, utilizing the current bandwidth resources sufficiently. In addition, MPLS TE can reserve resources during the creation of LSP tunnels to ensure the QoS.

Procedure

Step 1 Enable MPLS TE and RSVP-TE.

1. In the global config mode, run the **mpls** command to enter the MPLS mode.
2. In the MPLS mode, run the **mpls te** command to enable global MPLS TE, run the **mpls rsvp-te** command to enable global RSVP-TE, and run the **mpls te cspf** command to enable Constraint Shortest Path First (CSPF).
3. Run the **quit** command to quit the MPLS mode and run the **interface vlanif** command to enter the VLAN interface mode.
4. In the VLAN interface mode, run the **mpls** command to enable the VLAN interface MPLS, run the **mpls te** command to enable the VLAN interface MPLS TE, and run the **mpls rsvp-te** command to enable the VLAN interface RSVP-TE.

   NOTE

   - CSPF provides a way to select the path in an MPLS area. Enable CSPF before configuring other CSPF functions.
   - It is recommended that you configure CSPF on all transit nodes lest the ingress cannot calculate the entire path.

Step 2 (Optional) Configure the line bandwidth.

To guarantee the bandwidth of the service transmitted on the MPLS TE tunnel, perform this operation.

1. In the VLAN interface mode, run the **mpls te bandwidth max-reservable-bandwidth** command to configure the maximum reservable bandwidth for the MPLS TE tunnel on the VLAN interface.
2. In the VLAN interface mode, run the **mpls te bandwidth { bc0 bandwidth | bc1 bandwidth }** command to configure the bandwidth that can be obtained from BC0 and BC1 of the VLAN interface when an MPLS TE tunnel is created.

   NOTE

   - BC0: Indicates the global pool bandwidth of an MPLS TE tunnel.
   - BC1: Indicates the sub-pool bandwidth type of an MPLS TE tunnel. It is used to transmit services with higher priority and higher performance requirements.
   - The bandwidth values must meet the following requirement: maximum reservable bandwidth ≥ BC0 bandwidth ≥ BC1 bandwidth.

Step 3 Enable MPLS TE for the OSPF area.

The MA5800 enables the MPLS TE to know the relevant dynamic TE attributes of each link by extending the OSPF protocol. The extended OSPF enables the link status entry to add TE attributes, such as link bandwidth and affinity attribute. Each router in the network collects all the TE information in OSPF area and generates traffic engineering database (TEDB).

1. In the global config mode, run the **ospf** command to start the OSPF process and enter the OSPF mode.

 Copyright © Huawei Technologies Co., Ltd.

2. Run the **opaque-capability enable** command to enable the OSPF opaque capability.

   After the opaque capability of the MA5800 is enabled, it can export TEDB information to neighbor devices.

3. Run the **area** command to enter the OSPF area mode and run the **mpls-te enable** command to enable the OSPF area TE.

Step 4 (Optional) Configure an MPLS TE explicit path.

An explicit path consists of a series of nodes, which constitute a vector path according to the configured sequence. The IP address in an explicit path is the IP address of the interface on the node. Generally, the loopback interface IP address on the egress is used as the destination IP address of the explicit path.

To specify a known path for a special traffic stream in the MPLS network, you can run the **explicit-path** command in the global config mode to configure an explicit path, and then run the **mpls te path explicit-path** command in the tunnel mode to specify the explicit path for the tunnel.

After an explicit path is created, you can run the **next hop**, **modify hop**, and **delete hop** command to add a next hop node, modify a node, and delete a node respectively for the explicit path.

Step 5 Configure an MPLS TE tunnel interface.

1. In the global config mode, run the **interface tunnel** command to create a tunnel interface and enter the tunnel interface mode.
2. Run the **tunnel-protocol mpls te** command to configure the tunnel protocol to MPLS TE.
3. Run the **destination ip-address** command to configure the destination IP address of the tunnel. Generally, the egress LSR ID is used.
4. Run the **mpls te tunnel-id** command to configure the tunnel ID.
5. Run the **mpls te signal-protocol rsvp-te** command to configure the signaling protocol of the tunnel to RSVP-TE.
6. (Optional) Run the **mpls te bandwidth** command to configure the bandwidth for the tunnel. After the configuration is completed, only the VLAN interface that meets this bandwidth value can be selected as the node traversed by the MPLS TE tunnel path when the MPLS TE tunnel is created.

   If the MPLS TE tunnel is only used to change the data transmission path, you may not configure the tunnel bandwidth.

7. (Optional) Run the **mpls te path explicit-path** command to configure the explicit path used by the MPLS TE tunnel.

   If only the bandwidth used by the MPLS TE tunnel is limited but the transmission path is not limited, you may not configure the explicit path used by the MPLS TE tunnel.

Step 6 Check the configuration.

1. Run the **display mpls te cspf tedb** command to query the CSPF TEDB information.
2. Run the **display mpls te link-administration admission-control** command to check the CR LSP information allowed on the link, including the bandwidth and priority.

 Copyright © Huawei Technologies Co., Ltd.

3. Run the **display mpls te tunnel** command to query details about a specified tunnel.
4. Run the **display mpls te tunnel path** command to query the path information about a tunnel on a local node.
5. Run the **display mpls te tunnel-interface** command to query the tunnel interface information about a local node.

----End

Example

To configure the RSVP-TE LSP from the MA5800 to the PTN, set the parameters as follows.

- Set the parameters on the MA5800.
  - LSR ID: 10.3.3.3
  - Layer 3 interface IP address of VLAN 20 for MPLS forwarding: 10.1.1.3/24
  - Maximum reservable bandwidth of the VLAN interface: 20480 kbit/s; BC0 bandwidth: 10240 kbit/s
  - OSPF process ID: 100; OSPF area ID: 1
  - MPLS TE tunnel ID: 10; tunnel interface ID: 10
  - Required BC0 bandwidth when an MPLS TE tunnel is created: 5120 kbit/s
  - Other parameters: default settings
- Set the LSR ID of the PTN to 10.5.5.5.

```
huawei(config)#interface loopback 0
huawei(config-if-loopback0)#ip address 10.3.3.3 32
huawei(config-if-loopback0)#quit
huawei(config)#mpls lsr-id 10.3.3.3
huawei(config)#mpls
huawei(config-mpls)#mpls te
huawei(config-mpls)#mpls rsvp-te
//Configure the MPLS TE to use CSPF to calculate the shortest path to a node. huawei(config-mpls)#mpls te cspf
huawei(config-mpls)#quit

huawei(config)#interface vlanif 20
//Configure the IP address of the VLAN interface.
huawei(config-if-vlanif20)#ip address 10.1.1.3 24
//Enable MPLS for the VLAN interface.
huawei(config-if-vlanif20)#mpls
//Enable MPLS TE for the VLAN interface.
huawei(config-if-vlanif20)#mpls te
//Enable MPLS RSVP-TE for the VLAN interface.
huawei(config-if-vlanif20)#mpls rsvp-te
huawei(config-if-vlanif20)#quit
huawei(config)#ospf 100
//Enable the opaque capability to send the engineering data base information to peripheral devices.
huawei(config-ospf-100)#opaque-capability enable
huawei(config-ospf-100)#area 1
//Enable MPLS TE for the OSPF area.

huawei(config-ospf-100-area-0.0.0.1)#mpls-te enable standard-complying
huawei(config-ospf-100-area-0.0.0.1)#quit
huawei(config-ospf-100)#quit
huawei(config)#interface vlanif 20
//Configure the maximum reservable bandwidth of the Layer 3 interface.
huawei(config-if-vlanif20)#mpls te bandwidth max-reservable-bandwidth 20480
//Configure the obtainable maximum bandwidth of the Layer 3 interface from BC0 when the MPLS TE tunnel is created.
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-if-vlanif20)#mpls te bandwidth bc0 10240
 huawei(config-if-vlanif20)#quit
huawei(config)#interface tunnel 10
//Configure the link layer encapsulation protocol to MPLS TE for the tunnel interface, that is, configure the tunnel interface to work in the CR-LSP tunnel mode.
huawei(config-if-tunnel10)#tunnel-protocol mpls te
//Configure the destination IP address of the MPLS TE tunnel.
huawei(config-if-tunnel10)#destination 10.3.3.3
//Configure the MPLS TE tunnel ID, which, along with the LSR-ID, uniquely indicates an MPLS TE tunnel.
huawei(config-if-tunnel10)#mpls te tunnel-id 10
//Configure the protocol of the MPLS TE tunnel to RSVP-TE.
huawei(config-if-tunnel10)#mpls te signal-protocol rsvp-te
//Configure the global pool bandwidth required by the MPLS TE tunnel.

huawei(config-if-tunnel10)#mpls te bandwidth ct0 5120
//Allow the MPLS TE tunnel to be bound to a VPN instance, that is, the MPLS TE tunnel can function as the outer tunnel of the PWE3 service.

huawei(config-if-tunnel10)#mpls te reserved-for-binding
huawei(config-if-tunnel10)#quit
```

6.7.4 Configuring the MPLS RSVP-TE FRR

The RSVP TE FRR technology is applied to the MPLS TE network for implementing partial network protection. Specifically, when a certain link or node in the network fails, the LSP configured with FRR can automatically switch the data to the protect link.

Prerequisites

1. The IP address of the loopback interface must be configured.
2. The LSR ID must be configured.
3. The VLAN for MPLS label forwarding must be created.
4. Global MPLS and VLAN MPLS must be enabled.
5. The OSPF protocol must be successfully configured on each device in the network (the host route of each port must be successfully advertised).

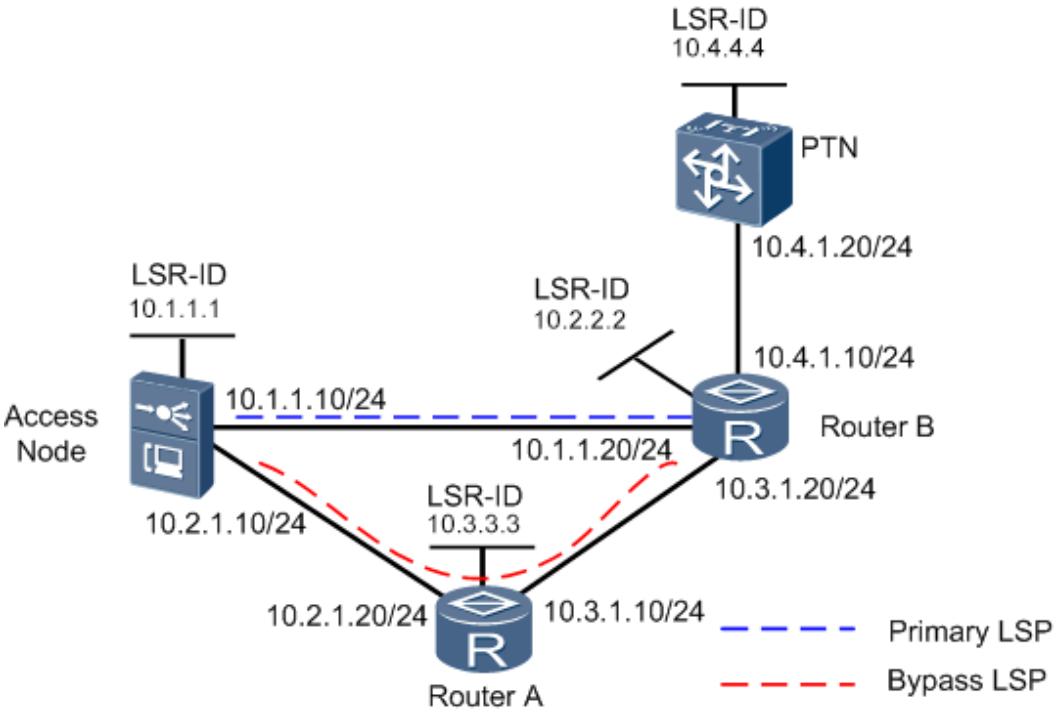
Context

- The implementation of the FRR is based on the extended RSVP-TE signaling. For the FRR, a protect tunnel is created in advance to protect the working tunnel. This prevents the broadcast delay of the notification between NEs and the duration for re-selecting the tunnel if the working tunnel fails. Therefore, the FRR can implement the second-level protection switchover.
- The MA5800 adopts the bypass mode (that is, using a protect path to protect multiple LSPs; the protect path is called the bypass LSP). **Figure 6-12** shows the FRR function implemented in the bypass mode.

As shown in the figure, the blue dotted line indicates the primary LSP and the red dotted line indicates the bypass LSP. When the link or node between the MA5800 and Router B is faulty, services are switched to the bypass link MA5800->Router A->Router B. In this manner, the LSP is not affected by the link or node fault.

 Copyright © Huawei Technologies Co., Ltd.

Figure 6-12 Example network of the MPLS RSVP-TE FRR



Procedure

Step 1 Enable MPLS TE and RSVP-TE.

1. In the global config mode, run the **mpls** command to enter the MPLS mode.
2. In the MPLS mode, run the **mpls te** command to enable global MPLS TE, run the **mpls rsvp-te** command to enable global RSVP-TE, and run the **mpls te cspf** command to enable Constraint Shortest Path First (CSPF).
3. Run the **quit** command to quit the MPLS mode and run the **interface vlanif** command to enter the VLAN interface mode.
4. In the VLAN interface mode, run the **mpls** command to enable the VLAN interface MPLS, run the **mpls te** command to enable the VLAN interface MPLS TE, and run the **mpls rsvp-te** command to enable the VLAN interface RSVP-TE.

   NOTE

   - CSPF provides a way to select the path in an MPLS area. Enable CSPF before configuring other CSPF functions.
   - It is recommended that you configure CSPF on all transit nodes lest the ingress cannot calculate the entire path.

Step 2 (Optional) Configure the line bandwidth.

To guarantee the bandwidth of the service transmitted on the MPLS TE tunnel, perform this operation.

1. In the VLAN interface mode, run the **mpls te bandwidth max-reservable-bandwidth** command to configure the maximum reservable bandwidth for the MPLS TE tunnel on the VLAN interface.

 Copyright © Huawei Technologies Co., Ltd.

2. In the VLAN interface mode, run the **mpls te bandwidth { bc0 bandwidth | bc1 bandwidth }** command to configure the bandwidth that can be obtained from BC0 and BC1 of the VLAN interface when an MPLS TE tunnel is created.

   NOTE

   - BC0: Indicates the global pool bandwidth of an MPLS TE tunnel.
   - BC1: Indicates the sub-pool bandwidth type of an MPLS TE tunnel. It is used to transmit services with higher priority and higher performance requirements.
   - The bandwidth values must meet the following requirement: maximum reservable bandwidth ≥ BC0 bandwidth ≥ BC1 bandwidth.

Step 3 Enable MPLS TE for the OSPF area.

The MA5800 enables the MPLS TE to know the relevant dynamic TE attributes of each link by extending the OSPF protocol. The extended OSPF enables the link status entry to add TE attributes, such as link bandwidth and affinity attribute. Each router in the network collects all the TE information in OSPF area and generates traffic engineering database (TEDB).

1. In the global config mode, run the **ospf** command to start the OSPF process and enter the OSPF mode.
2. Run the **opaque-capability enable** command to enable the OSPF opaque capability.

   After the opaque capability of the MA5800 is enabled, it can export TEDB information to neighbor devices.
3. Run the **area** command to enter the OSPF area mode and run the **mpls-te enable** command to enable the OSPF area TE.

Step 4 Set up the primary tunnel on the MA5800.

1. Configure the explicit path of the primary LSP.

   An explicit path consists of a series of nodes, which constitute a vector path according to the configured sequence. The IP address in an explicit path is the IP address of the interface on the node. Generally, the loopback interface IP address on the egress is used as the destination IP address of the explicit path.

   To specify a known path for a special traffic stream in the MPLS network, you can run the **explicit-path** command in the global config mode to configure an explicit path, and then run the **mpls te path explicit-path** command in the tunnel mode to specify the explicit path for the tunnel.

   After an explicit path is created, you can run the **next hop**, **modify hop**, and **delete hop** command to add a next hop node, modify a node, and delete a node respectively for the explicit path.
2. Configure the MPLS TE tunnel of the primary LSP.
   a. In the global config mode, run the **interface tunnel** command to create a tunnel interface and enter the tunnel interface mode.
   b. Run the **tunnel-protocol mpls te** command to configure the tunnel protocol to MPLS TE.
   c. Run the **destination ip-address** command to configure the destination IP address of the tunnel. Generally, the egress LSR ID is used.
   d. Run the **mpls te tunnel-id** command to configure the tunnel ID.
   e. Run the **mpls te signal-protocol rsvp-te** command to configure the signaling protocol of the tunnel to RSVP-TE.

 Copyright © Huawei Technologies Co., Ltd.

f. (Optional) Run the **mpls te bandwidth** command to configure the bandwidth for the tunnel. After the configuration is completed, only the VLAN interface that meets this bandwidth value can be selected as the node traversed by the MPLS TE tunnel path when the MPLS TE tunnel is created.

If the MPLS TE tunnel is only used to change the data transmission path, you may not configure the tunnel bandwidth.

g. Run the **mpls te path explicit-path** command to configure the explicit path used by the MPLS TE tunnel.

3. Enable the FRR function of the tunnel.

Run the **mpls te fast-reroute [bandwidth]** command to enable TE FRR of the tunnel interface and allow bandwidth protection. By default, the FRR function is prohibited.

NOTE

Bandwidth protection configured through this command is used only for selecting the bypass tunnel policy. When the primary tunnel is faulty and needs to switch to a bypass tunnel, the bypass tunnel that meets the bandwidth requirement is preferred. If no bypass tunnel meets the bandwidth requirement, the primary tunnel selects an optimal bypass tunnel from the existing bypass tunnels.

Step 5 Set up a bypass LSP tunnel on the MA5800.

1. Configure the explicit path of the bypass LSP.
   a. In the global config mode, run the **explicit-path** command to configure the explicit path. In the tunnel mode, run the **mpls te path explicit-path** command to specify the explicit path for the tunnel.
   b. Run the **next hop**, **modify hop**, and **delete hop** command to add a next hop node, modify a node, and delete a node respectively for the explicit path.
2. Configure the MPLS TE tunnel of the bypass LSP.

NOTICE

MPLS TE tunnel IDs of the primary and bypass LSPs cannot be the same.

   a. In the global config mode, run the **interface tunnel** command to create a tunnel interface and enter the tunnel interface mode.
   b. Run the **tunnel-protocol mpls te** command to configure the tunnel protocol to MPLS TE.
   c. Run the **destination ip-address** command to configure the destination IP address of the tunnel. Generally, the egress LSR ID is used.
   d. Run the **mpls te tunnel-id** command to configure the tunnel ID.
   e. Run the **mpls te signal-protocol rsvp-te** command to configure the signaling protocol of the tunnel to RSVP-TE.
   f. (Optional) Run the **mpls te bandwidth** command to configure the bandwidth for the tunnel. After the configuration is completed, only the VLAN interface that meets this bandwidth value can be selected as the node traversed by the MPLS TE tunnel path when the MPLS TE tunnel is created.

 Copyright © Huawei Technologies Co., Ltd.

If the MPLS TE tunnel is only used to change the data transmission path, you may not configure the tunnel bandwidth.

g. Run the **mpls te path explicit-path** command to configure the explicit path used by the MPLS TE tunnel.

3. Bind the bypass LSP tunnel to the protected interface.

a. In the tunnel mode, run the **mpls te bypass-tunnel** command to configure a bypass tunnel of the FRR.

NOTE

The total bandwidth of all LSPs that use bypass tunnels does not exceed the bandwidth of the primary tunnel. If multiple bypass tunnels exist, the system uses the best-fit algorithm to determine which bypass to use.

b. Run the **mpls te protected-interface** command to specify the interface to be protected by the bypass tunnel. When the interface is faulty, a bypass tunnel switching is triggered.

NOTE

One bypass tunnel can protect up to three interfaces, and MPLS TE must be enabled for the protected interfaces.

----End

Result

Enter the VLAN interface mode, and run the **shutdown** command to shut down the VLAN interface to disable the protected egress on the primary LSP. Then run the **display interface tunnel** command to query the status of the primary LSP on the MA5800. You can see that the tunnel interface is still in the up state. Finally, run the **tracert lsp te tunnel** command to check the path traversed by the tunnel. You can see that the link is switched to the bypass tunnel.

Example

As shown in **Figure 6-12**, when the link or node between the MA5800 and Router B is faulty, services are switched to the standby link MA5800->Router A->Router B. In this manner, the LSP is not affected by the fault of link or node. Set the parameters as follows:

- Set the parameters on the MA5800.
  - LSR ID: 10.1.1.1
  - IP address of VLAN interface 10 connected to Router B: 10.1.1.10/24
  - IP address of VLAN interface 20 connected to Router A: 10.2.1.10/24
- Set the parameters on the Router B.
  - LSR ID: 10.2.2.2
  - IP address of the interface connected to the MA5800: 10.1.1.20/24
  - IP address of the interface connected to Router A: 10.3.1.20/24
  - IP address of the interface connected to the PTN: 10.4.1.10/24
- Set the parameters on the Router A.
  - LSR ID: 10.3.3.3

 Copyright © Huawei Technologies Co., Ltd.

  - IP address of the interface connected to the MA5800: 10.2.1.20/24
  - IP address of the interface connected to Router B: 10.3.1.10/24
- Set the parameters on the PTN.
  - LSR ID: 10.4.4.4
  - IP address of the interface connected to Router B: 10.4.1.20/24

```
  //Configure the LSR-ID.
huawei(config)#interface loopback 0
huawei(config-if-loopback0)#ip address 10.1.1.1 32
huawei(config-if-loopback0)#quit
huawei(config)#mpls lsr-id 10.1.1.1
  //Enable RSVP-TE.
huawei(config)#mpls
huawei(config-mpls)#mpls te
huawei(config-mpls)#mpls rsvp-te
huawei(config-mpls)#mpls te cspf
huawei(config-mpls)#quit
  //Configure the IP address of VLAN interface 10 and enable RSVP-TE of the VLAN interface.
huawei(config)#vlan 10 standard
huawei(config)#interface vlanif 10
huawei(config-if-vlanif10)#ip address 10.1.1.10 24
huawei(config-if-vlanif10)#mpls
huawei(config-if-vlanif10)#mpls te
huawei(config-if-vlanif10)#mpls rsvp-te
huawei(config-if-vlanif10)#quit
  //Configure the IP address of VLAN interface 20 and enable RSVP-TE of the VLAN interface.
huawei(config)#vlan 20 standard
huawei(config)#interface vlanif 20
huawei(config-if-vlanif20)#ip address 10.2.1.10 24
huawei(config-if-vlanif20)#mpls
huawei(config-if-vlanif20)#mpls te
huawei(config-if-vlanif20)#mpls rsvp-te
huawei(config-if-vlanif20)#quit
  //Configure OSPF TE.
huawei(config)#ospf 100
huawei(config-ospf-100)#opaque-capability enable
huawei(config-ospf-100)#area 0
huawei(config-ospf-100-area-0.0.0.0)#mpls-te enable standard-complying
huawei(config-ospf-100-area-0.0.0.0)#quit
huawei(config-ospf-100)#quit
  //Configure the explicit path of the primary LSP.

huawei(config)#explicit-path pri-path
huawei(config-explicit-path-pri-path)#next hop 10.1.1.20
huawei(config-explicit-path-pri-path)#next hop 10.4.1.20
huawei(config-explicit-path-pri-path)#quit
  //Configure the MPLS TE tunnel of the primary LSP.
huawei(config)#interface tunnel 10
huawei(config-if-tunnel10)#tunnel-protocol mpls te
huawei(config-if-tunnel10)#destination 10.2.2.2
huawei(config-if-tunnel10)#mpls te tunnel-id 10
huawei(config-if-tunnel10)#mpls te signal-protocol rsvp-te
huawei(config-if-tunnel10)#mpls te path explicit-path pri-path
huawei(config-if-tunnel10)#mpls te fast-reroute
huawei(config-if-tunnel10)#mpls te commit
huawei(config-if-tunnel10)#quit
  //Configure the explicit path of the bypass LSP.

huawei(config)#explicit-path bypass-path
huawei(config-explicit-path-bypass-path)#next hop 10.2.1.20
huawei(config-explicit-path-bypass-path)#next hop 10.3.1.20
huawei(config-explicit-path-bypass-path)#quit
  //Configure the MPLS TE tunnel of the bypass LSP.
huawei(config)#interface tunnel 20
huawei(config-if-tunnel20)#tunnel-protocol mpls te
huawei(config-if-tunnel20)#destination 10.2.2.2
huawei(config-if-tunnel20)#mpls te tunnel-id 20
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-if-tunnel20)#mpls te signal-protocol rsvp-te
huawei(config-if-tunnel20)#mpls te path explicit-path bypass-path
huawei(config-if-tunnel20)#mpls te bypass-tunnel
huawei(config-if-tunnel20)#mpls te protected-interface vlanif 10
huawei(config-if-tunnel20)#mpls te commit
huawei(config-if-tunnel20)#quit
```

6.7.5 Configuring the MPLS OAM

The MPLS OAM function uses an effective OAM mechanism to detect whether an LSP is normal and report an alarm in time when an LSP fault occurs. In addition, the MPLS OAM function features a complete protection switching mechanism, which triggers a switchover when a defect at the MPLS layer is detected to minimize the data loss.

Context

Through the MPLS OAM mechanism, the MA5800 can effectively detect, confirm, and locate internal defects at the MPLS layer of a network. Then, the system reports and handles the defects. In addition, the system provides a mechanism for triggering 1:1 protection switching when a fault occurs.

The basic process of the MPLS OAM connectivity check and protection switching is as follows:

1. The source transmits the CV/FFD packets to the destination through the detected LSP.
2. The destination checks the correctness of the type and frequency carried in the received detection packets and measures the number of correct and incorrect packets that are received within the detection period to monitor the connectivity of the LSP in real time.
3. After detecting a defect, the destination transmits the BDI packets that carry the defect information to the source through the backward path.
4. The source learns about the status of the defect, and triggers the corresponding protection switching when the protect group is correctly configured.

Configure the MPLS OAM as follows:

1. Configure the active LSP at the source end (ingress).
2. Configure the standby LSP at the source end.
3. Create a tunnel protect group.
4. Enable the MPLS OAM function at the source end.
5. Configure the backward LSP at the destination end (egress).
6. Enable the MPLS OAM function at the destination end.

NOTE

If only the MPLS OAM connectivity check needs to be enabled and 1:1 protection is not required for the LSP, you need not configure the standby LSP or the tunnel protect group at the source end.

 Copyright © Huawei Technologies Co., Ltd.

6.7.5.1 Configuration Example for Detection of MPLS OAM for Static LSP Connectivity

This topic describes how to configure the function of MPLS OAM to detect the static LSP connectivity.

Prerequisites

Before the configuration, make sure that:

- Set the IP addresses and the masks of the ports based on the example network. After that, LSRs can ping the peer LSRs.
- A static or dynamic route must be successfully configured on each device in the network (so that LSRs can reach each other through the IP route).

Networking

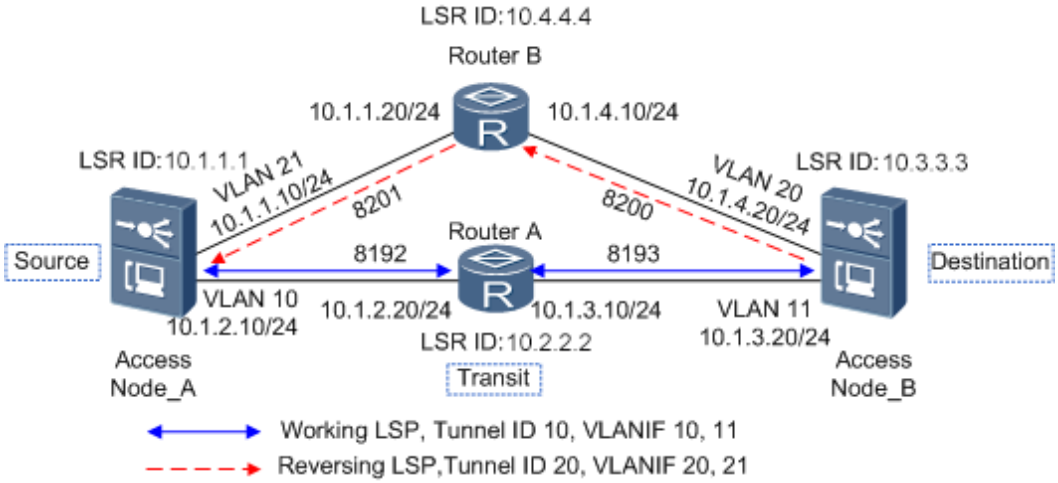
Figure 6-13 shows an example network of configuring MPLS OAM to detect the static LSP connectivity.

1. Source end MA5800_A sends CV/FFD detection packets to the destination end through the detected LSP (MA5800_A->Router A->MA5800_B).
2. After detecting a defect, the destination transmits the BDI packets that carry the defect information to the source through the backward LSP (MA5800_B->Router B->MA5800_A). This enables the source end to obtain the defect status in time.

NOTE

To facilitate description of the MPLS OAM application, the MA5800 is used at both the source end and destination end as an example. In the actual application, the MA5800 at one end may be replaced by a device that supports MPLS OAM such as a PTN device, but their implementation principles are the same.

Figure 6-13 Example network of detection of MPLS OAM for static LSP connectivity



 Copyright © Huawei Technologies Co., Ltd.

Data Plan

Table 6-5 provides the data plan for detection of MPLS OAM for static LSP connectivity.

Table 6-5 Data plan for detection of MPLS OAM for static LSP connectivity

Item	Data
MA5800_A	LSR ID: 10.1.1.1
	Port: 0/9/0 IP address of VLAN interface 10 connected to Router A: 10.1.2.10/24 Tunnel ID: 10; tunnel interface ID: 10 Out label value of the LSP ingress: 8192 In label value of the LSP egress: 8193
	Port: 0/9/1 IP address of VLAN interface 21 connected to Router B: 10.1.1.10/24
	Static LSP: Router A to MA5800_B
MA5800_B	LSR ID: 10.3.3.3
	Port: 0/9/0 IP address of VLAN interface 11 connected to Router A: 10.1.3.20/24
	Port: 0/9/1 IP address of VLAN interface 20 connected to Router B: 10.1.4.20/24 Tunnel ID: 20; tunnel interface ID: 20 Out label value of the LSP ingress: 8200 In label value of the LSP egress: 8201
	Static LSP: Router B to MA5800_A
Router A	LSR ID: 10.2.2.2
	IP address of the interface connected to the MA5800_A: 10.1.2.20/24
	IP address of the interface connected to the MA5800_B: 10.1.3.10/24
Router B	LSR ID: 10.4.4.4
	IP address of the interface connected to the MA5800_A: 10.1.1.20/24
	IP address of the interface connected to the MA5800_B: 10.1.4.10/24

 Copyright © Huawei Technologies Co., Ltd.

Procedure

- **Configure source end MA5800_A.**
  a. Configure the loopback interface.

```
huawei(config)#interface loopback 0
huawei(config-if-loopback0)#ip address 10.1.1.1 32
huawei(config-if-loopback0)#quit
```

  b. Enable the basic MPLS and MPLS TE.

     i. Enable the basic MPLS and MPLS TE globally.

```
huawei(config)#mpls lsr-id 10.1.1.1
huawei(config)#mpls
huawei(config-mpls)#mpls te
huawei(config-mpls)#quit
```

     ii. Enable the basic MPLS and MPLS TE on the interface.

```
huawei(config)#vlan 10 standard
huawei(config)#port vlan 10 0/9 0
huawei(config)#interface vlanif 10
huawei(config-if-vlanif10)#ip address 10.1.2.10 24
huawei(config-if-vlanif10)#mpls
huawei(config-if-vlanif10)#mpls te
huawei(config-if-vlanif10)#quit
huawei(config)#vlan 21 standard
huawei(config)#port vlan 21 0/9 1
huawei(config)#interface vlanif 21
huawei(config-if-vlanif21)#ip address 10.1.1.10 24
huawei(config-if-vlanif21)#mpls
huawei(config-if-vlanif21)#mpls te
huawei(config-if-vlanif21)#quit
```

  c. Configure the MPLS TE tunnel from the source end to the destination end.

     Configure the MPLS TE tunnel bound to the detected LSP.

```
huawei(config)#interface tunnel 10
huawei(config-if-tunnel10)#tunnel-protocol mpls te
huawei(config-if-tunnel10)#destination 10.3.3.3
huawei(config-if-tunnel10)#mpls te tunnel-id 20
huawei(config-if-tunnel10)#mpls te signal-protocol static
huawei(config-if-tunnel10)#mpls te commit
huawei(config-if-tunnel10)#quit
```

  d. Configure the static LSP bound to the MPLS TE tunnel.

     Source end MA5800 functions as the ingress of the detected static LSP.

```
huawei(config)#static-lsp ingress tunnel-interface tunnel 10
destination 10.3.3.3 nexthop 10.1.2.20 out-label 8192
```

     Source end MA5800 functions as the egress of the detected static LSP.

```
huawei(config)#static-lsp egress LSP1 incoming-interface vlanif 10 in-label 8193
```

     Source end MA5800 functions as the egress of the backward static LSP.

```
huawei(config)#static-lsp egress LSP2 incoming-interface vlanif 20 in-label 8201
```

  e. Enable MPLS OAM at source end MA5800_A.

```
huawei(config)#mpls
huawei(config-mpls)#mpls oam
huawei(config-mpls)#quit
huawei(config)#mpls oam ingress tunnel 10 type ffd frequency 100
backward-lsp lsr-id 10.3.3.3 tunnel-id 20
...//Configure the MPLS OAM source end. Configure the tunnel ID of the detected LSP to 10,
detection packet type to FFD, Tx frequency to 100 ms,  LSR-ID of the backward LSP to 10.3.3.3,
```

 Copyright © Huawei Technologies Co., Ltd.

```
...//and backward LSP tunnel ID to 20.
huawei(config)#mpls oam ingress enable all
```

f. Save the data.

```
huawei(config)#save
```

- **Configure Router A or Router B.**

  When functioning as the transit node, Router A or Router B mainly forwards MPLS labels. The ingress interface, in label, next hop IP address, and out label must be configured bi-directionally. For detailed configuration, see the configuration guide of the specific router.

- **Configure destination end MA5800_B.**

  a. Configure the loopback interface.

  ```
  huawei(config)#interface loopback 0
  huawei(config-if-loopback0)#ip address 10.3.3.3 32
  huawei(config-if-loopback0)#quit
  ```

  b. Enable the basic MPLS and MPLS TE.

  i. Enable the basic MPLS and MPLS TE globally.

  ```
  huawei(config)#mpls lsr-id 10.3.3.3
  huawei(config)#mpls
  huawei(config-mpls)#mpls te
  huawei(config-mpls)#quit
  ```

  ii. Enable the basic MPLS and MPLS TE on the interface.

  ```
  huawei(config)#vlan 11 standard
  huawei(config)#port vlan 11 0/9 0
  huawei(config)#interface vlanif 11
  huawei(config-if-vlanif11)#ip address 10.1.3.20 24
  huawei(config-if-vlanif11)#mpls
  huawei(config-if-vlanif11)#mpls te
  huawei(config-if-vlanif11)#quit
  huawei(config)#vlan 20 standard
  huawei(config)#port vlan 20 0/9 1
  huawei(config)#interface vlanif 20
  huawei(config-if-vlanif20)#ip address 10.1.4.20 24
  huawei(config-if-vlanif20)#mpls
  huawei(config-if-vlanif20)#mpls te
  huawei(config-if-vlanif20)#quit
  ```

  c. Configure the MPLS TE tunnel from the destination end to the source end.

  Configure the MPLS TE tunnel bound to the detected LSP.

  ```
  huawei(config)#interface tunnel 10
  huawei(config-if-tunnel10)#tunnel-protocol mpls te
  huawei(config-if-tunnel10)#destination 10.1.1.1
  huawei(config-if-tunnel10)#mpls te tunnel-id 10
  huawei(config-if-tunnel10)#mpls te signal-protocol static
  huawei(config-if-tunnel10)#mpls te commit
  huawei(config-if-tunnel100)#quit
  ```

  Configure the MPLS TE tunnel bound to the backward LSP.

  ```
  huawei(config)#interface tunnel 20
  huawei(config-if-tunnel20)#tunnel-protocol mpls te
  huawei(config-if-tunnel20)#destination 10.1.1.1
  huawei(config-if-tunnel20)#mpls te tunnel-id 20
  huawei(config-if-tunnel20)#mpls te signal-protocol static
  huawei(config-if-tunnel20)#mpls te commit
  huawei(config-if-tunnel20)#quit
  ```

  d. Configure the static LSP bound to the tunnel.

  Destination end MA5800 functions as the egress of the detected static LSP.

  ```
  huawei(config)#static-lsp egress LSP2 incoming-interface vlanif 10 in-label 8192
  ```

 Copyright © Huawei Technologies Co., Ltd.

Destination end MA5800 functions as the ingress of the detected static LSP.

```
huawei(config)#static-lsp ingress tunnel-interface tunnel 10
destination 10.1.1.1 nexthop 10.1.3.10 out-label 8193
```

Destination end MA5800 functions as the ingress of the backward static LSP.

```
huawei(config)#static-lsp ingress tunnel-interface tunnel 20
destination 10.1.1.1 nexthop 10.1.4.10 out-label 8200
```

e. Enable MPLS OAM at destination end MA5800.

```
huawei(config)#mpls
huawei(config-mpls)#mpls oam
huawei(config-mpls)#quit
huawei(config)#mpls oam egress lsr-id 10.1.1.1 tunnel-id 10 type ffd frequency 100
backward-lsp t
unnel 20 private
...//Configure the MPLS OAM destination end. Configure the ingress LSR-ID of the detected LSP to 10.1.1.1, tunnel ID to 10, detection packet type to FFD, Tx frequency to 100 ms,

...//backward LSP tunnel ID to 20, and tunnel to exclusive mode.

huawei(config)#mpls oam egress enable all
```

f. Save the data.

```
huawei(config)#save
```

----End

Result

After the configuration, shut down the interface of VLAN 10 by running the **shutdown** command on MA5800_A to simulate the link fault:

- On MA5800_B, run the **display mpls oam egress** command and you can see the following defect state: dLocv detected (dLocv).
- On MA5800_A, run the **display mpls oam ingress** command and you can see the following defect state: in defect (In-defect).

Perform similar operations on MA5800_B and you can obtain similar results.

6.7.5.2 Configuration Example of the MPLS OAM Protection Switching Function

This topic describes how to configure MPLS OAM to implement the protection switching function.

Service Requirements

- The OAM mechanism is used to detect in real time whether the MPLS link is normal and generates an alarm in time when a link fault is detected.
- The end-to-end tunnel protection technology is provided to recover the interrupted service.
- RSVP-TE is used to create an LSP tunnel for the specified path and reserve resources so that the existing bandwidth resources can be fully used and QoS can be improved for specific services.

 Copyright © Huawei Technologies Co., Ltd.

Prerequisite

- The OSPF protocol must be successfully configured on each LSR in the network (the host route of each port must be successfully advertised).
- The interface IP address and mask, loopback interface, and LSR-ID must be configured on each LSR.
- The global and physical interface MPLS and MPLS TE functions must be enabled on each node of the LSR.

Networking

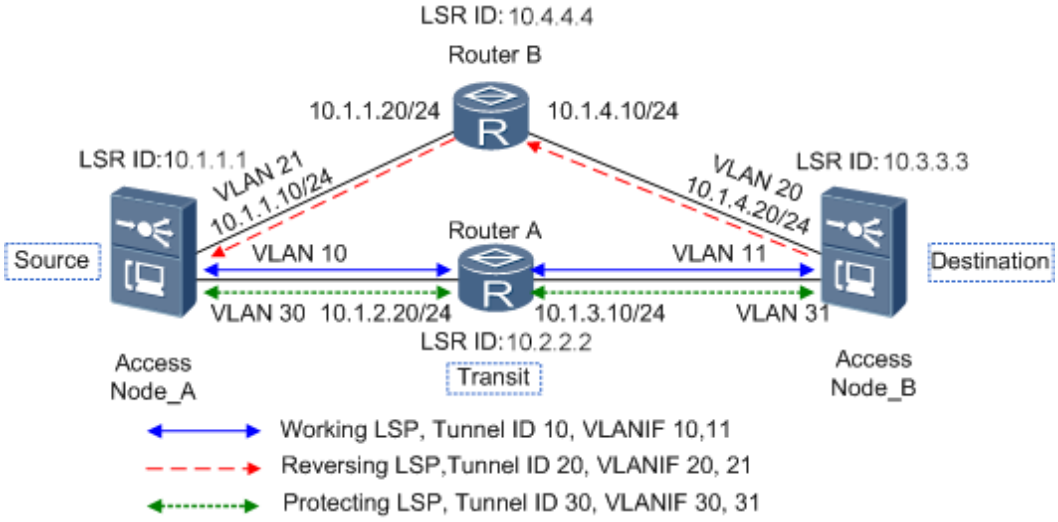
Figure 6-14 shows an example network for configuring the MPLS OAM protection switching function.

Configure two LSP tunnels on source end MA5800_A and destination end MA5800_B functioning primary and secondary LSPs. Enable the MPLS OAM protection switching function for the LSPs. When the primary LSP is faulty, the traffic is switched to the secondary LSP. Configure the backward LSP for reporting a fault to source end MA5800_A.

NOTE

To prevent a fault from occurring on a transit node (for example, router A), it is recommended that you specify different transit nodes when creating a secondary LSP.

Figure 6-14 Configuring the MPLS OAM protection switching function



Data Plan

Table 6-6 provides the data plan for the MPLS OAM protection switching.

Table 6-6 Data plan for the MPLS OAM protection switching

Item	Data
MA5800_A	LSR ID: 10.1.1.1

 Copyright © Huawei Technologies Co., Ltd.

Item	Data
	Port: 0/9/0 IP address of VLAN interface 10 connected to Router A: 10.1.2.10/24
	Port: 0/9/1 IP address of VLAN interface 30 connected to Router A: 10.1.5.10/24 IP address of VLAN interface 21 connected to Router B: 10.1.1.10/24
MA5800_B	LSR ID: 10.3.3.3
	Port: 0/9/0 IP address of VLAN interface 11 connected to Router A: 10.1.3.20/24
	Port: 0/9/1 IP address of VLAN interface 20 connected to Router B: 10.1.4.20/24 IP address of VLAN interface 31 connected to Router A: 10.1.6.20/24
	Backward tunnel: Router B to MA5800_A
Router A	LSR ID: 10.2.2.2
Router B	LSR ID: 10.4.4.4

Procedure

- **Configure source end MA5800_A.**
  a. Configure the loopback interface.

```
huawei(config)#interface loopback 0
huawei(config-if-loopback0)#ip address 10.1.1.1 32
huawei(config-if-loopback0)#quit
```

  b. Enable the basic MPLS, MPLS TE, and RSVP-TE functions.
     i. Enable the global basic MPLS, MPLS TE, and RSVP-TE functions.

```
huawei(config)#mpls lsr-id 10.1.1.1
huawei(config)#mpls
huawei(config-mpls)#mpls te
huawei(config-mpls)#mpls rsvp-te
huawei(config-mpls)#mpls te cspf
huawei(config-mpls)#quit
```

     ii. Enable the interface basic MPLS, MPLS TE, and RSVP-TE functions.

```
  //Configure the attributes of VLAN interface 10 and configure the L3 IP address of VLAN interface10 to 10.1.2.10/24.
huawei(config)#vlan 10 standard
huawei(config)#port vlan 10 0/9 0
huawei(config)#interface vlanif 10
huawei(config-if-vlanif10)#ip address 10.1.2.10 24
huawei(config-if-vlanif10)#mpls
huawei(config-if-vlanif10)#mpls te
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-if-vlanif10)#mpls rsvp-te
huawei(config-if-vlanif10)#mpls te bandwidth max-reservable-bandwidth 10240
  //(Optional) Configure VLAN interface 10 to provide a reservable bandwidth of 10240 kbit/s for all tunnels.
huawei(config-if-vlanif10)#quit
  //Configure the attributes of VLAN interface 30 and configure the L3 IP address of VLAN interface 30 to 10.1.5.10/24.
huawei(config)#vlan 30 standard
huawei(config)#port vlan 30 0/9 1
huawei(config)#interface vlanif 30
huawei(config-if-vlanif30)#ip address 10.1.5.10 24
huawei(config-if-vlanif30)#mpls
huawei(config-if-vlanif30)#mpls te
huawei(config-if-vlanif30)#mpls rsvp-te
huawei(config-if-vlanif30)#mpls te bandwidth max-reservable-bandwidth 10240
  //(Optional) Configure VLAN interface 30 to provide a reservable bandwidth of 10240 kbit/s for all tunnels.
huawei(config-if-vlanif30)#quit
  //Configure the attributes of VLAN interface 21 and configure the L3 IP address of VLAN interface 21 to 10.1.1.10/24.
huawei(config)#vlan 21 standard
huawei(config)#port vlan 21 0/9 1
huawei(config)#interface vlanif 21
huawei(config-if-vlanif21)#ip address 10.1.1.10 24
huawei(config-if-vlanif21)#mpls
huawei(config-if-vlanif21)#mpls te
huawei(config-if-vlanif21)#mpls rsvp-te
huawei(config-if-vlanif21)#mpls te bandwidth max-reservable-bandwidth 10240
  //(Optional) Configure VLAN interface 21 to provide a reservable bandwidth of 10240 kbit/s for all tunnels.
huawei(config-if-vlanif21)#quit
```

c. Enable MPLS TE for the OSPF area.

```
huawei(config)#ospf 100
huawei(config-ospf-100)#opaque-capability enable
huawei(config-ospf-100)#area 0
huawei(config-ospf-100-area-0.0.0.0)#mpls-te enable standard-complying
huawei(config-ospf-100-area-0.0.0.0)#quit
huawei(config-ospf-100)#quit
```

d. Configure the MPLS TE tunnel from the source end to the destination end.

Configure the attributes of the working MPLS TE tunnel from the source end to the destination end.

```
huawei(config)#interface tunnel 10
huawei(config-if-tunnel10)#tunnel-protocol mpls te
huawei(config-if-tunnel10)#destination 10.3.3.3
huawei(config-if-tunnel10)#mpls te tunnel-id 10
huawei(config-if-tunnel10)#mpls te signal-protocol rsvp-te
huawei(config-if-tunnel10)#mpls te bandwidth ct0 5120  //(Optional) Configure the global bandwidth of tunnel 10 to 5120 kbit/s.
huawei(config-if-tunnel10)#quit
```

Configure the attributes of the protection MPLS TE tunnel from the source end to the destination end.

```
huawei(config)#interface tunnel 30
huawei(config-if-tunnel30)#tunnel-protocol mpls te
huawei(config-if-tunnel30)#destination 10.3.3.3
huawei(config-if-tunnel30)#mpls te tunnel-id 30
huawei(config-if-tunnel30)#mpls te signal-protocol rsvp-te
huawei(config-if-tunnel30)#mpls te bandwidth ct0 5120  //(Optional) Configure the global bandwidth of tunnel 30 to 5120 kbit/s.
huawei(config-if-tunnel30)#quit
```

e. Configure a tunnel protect group.

 Copyright © Huawei Technologies Co., Ltd.

Configure tunnel 30 as the protect tunnel for tunnel 10, switching mode to revertive, and automatic WTR time to 900s (the corresponding WTR is 30 with step 30s).

```
huawei(config)#interface tunnel 10
huawei(config-if-tunnel10)#mpls te protection tunnel 30 mode revertive wtr 30
huawei(config-if-tunnel10)#quit
```

f. Enable MPLS OAM at source end MA5800_A.

```
huawei(config)#mpls
huawei(config-mpls)#mpls oam
huawei(config-mpls)#quit
huawei(config)#mpls oam ingress tunnel 10 type ffd frequency 100
backward-lsp lsr-id 10.3.3.3 tunnel-id 20
  //Configure the MPLS OAM source end. Configure the tunnel ID of the detected LSP to 10,
detection packet type to FFD, Tx frequency to 100 ms, LSR-ID of the backward LSP to 10.3.3.3,
  //and backward LSP tunnel ID to 20.
huawei(config)#mpls oam ingress enable all
```

g. Save the data.

```
huawei(config)#save
```

- **Configure Router A or Router B.**

  When functioning as the transit node, Router A or Router B mainly forwards MPLS labels. The ingress interface, in label, next hop IP address, and out label must be configured bi-directionally. For detailed configuration, see the configuration guide of the specific router.

- **Configure destination end MA5800_B.**

  a. Configure the loopback interface.

  ```
  huawei(config)#interface loopback 0
  huawei(config-if-loopback0)#ip address 10.3.3.3 32
  huawei(config-if-loopback0)#quit
  ```

  b. Enable the basic MPLS, MPLS TE, and RSVP-TE functions.

  i. Enable the global basic MPLS, MPLS TE, and RSVP-TE functions.

  ```
  huawei(config)#mpls lsr-id 10.3.3.3
  huawei(config)#mpls
  huawei(config-mpls)#mpls te
  huawei(config-mpls)#mpls rsvp-te
  huawei(config-mpls)#mpls te cspf
  huawei(config-mpls)#quit
  ```

  ii. Enable the interface basic MPLS, MPLS TE, and RSVP-TE functions.

  ```
    //Configure the attributes of VLAN interface 11 and configure the L3 IP address of VLAN
  interface 11 to 10.1.3.20/24.
  huawei(config)#vlan 11 standard
  huawei(config)#port vlan 11 0/9 0
  huawei(config)#interface vlanif 11
  huawei(config-if-vlanif11)#ip address 10.1.3.20 24
  huawei(config-if-vlanif11)#mpls
  huawei(config-if-vlanif11)#mpls te
  huawei(config-if-vlanif11)#mpls rsvp-te
  huawei(config-if-vlanif10)#quit
    //Configure the attributes of VLAN interface 20 and configure the L3 IP address of VLAN
  interface 20 to 10.1.4.20/24.
  huawei(config)#vlan 20 standard
  huawei(config)#port vlan 20 0/9 1
  huawei(config)#interface vlanif 20
  huawei(config-if-vlanif20)#ip address 10.1.4.20 24
  huawei(config-if-vlanif20)#mpls
  huawei(config-if-vlanif20)#mpls te
  huawei(config-if-vlanif20)#mpls rsvp-te
  huawei(config-if-vlanif20)#quit
    //Configure the attributes of VLAN interface 31 and configure the L3 IP address of VLAN
  interface 31 to 10.1.6.20/24.
  huawei(config)#vlan 31 standard
  ```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#port vlan 31 0/9 1
huawei(config)#interface vlanif 31
huawei(config-if-vlanif31)#ip address 10.1.6.20 24
huawei(config-if-vlanif31)#mpls
huawei(config-if-vlanif31)#mpls te
huawei(config-if-vlanif31)#mpls rsvp-te
huawei(config-if-vlanif31)#quit
```

c. Configure the MPLS TE tunnel bound to the backward LSP.

Configure the tunnel ID to 20, destination IP address to 10.1.1.1, and global bandwidth for the tunnel to 5120 kbit/s.

```
huawei(config)#interface tunnel 20
huawei(config-if-tunnel20)#tunnel-protocol mpls te
huawei(config-if-tunnel20)#destination 10.1.1.1
huawei(config-if-tunnel20)#mpls te tunnel-id 20
huawei(config-if-tunnel20)#mpls te signal-protocol rsvp-te
huawei(config-if-tunnel20)#mpls te bandwidth ct0 5120
huawei(config-if-tunnel20)#mpls te reserved-for-binding
huawei(config-if-tunnel20)#quit
```

d. Enable MPLS OAM at destination end MA5800_B.

```
huawei(config)#mpls
huawei(config-mpls)#mpls oam
huawei(config-mpls)#quit
huawei(config)#mpls oam egress lsr-id 10.1.1.1 tunnel-id 10 type ffd frequency 100 backward-lsp tunnel 20 private
  //Configure the MPLS OAM destination end. Configure the ingress LSR-ID of the detected LSP to 10.1.1.1, tunnel ID to 10, detection packet type to FFD, Tx frequency to 100 ms,
  //backward LSP tunnel ID to 20, and tunnel to exclusive mode.
huawei(config)#mpls oam egress enable all
```

e. Save the data.

```
huawei(config)#save
```

----End

Result

After the configuration, you can shut down the interface of VLAN 10 by running the **shutdown** command on MA5800_A to simulate the link fault. Then, you can query the information about the primary tunnel (with ID 10) that is configured on MA5800_A by running the **display mpls te protection tunnel** command on MA5800_A. The information is as follows:

- Status of the working tunnel (work-tunnel defect state): in defect.
- Status of the protection tunnel (protect-tunnel defect state): non-defect.
- Switch result: The traffic is switched to protection tunnel 30.

 Copyright © Huawei Technologies Co., Ltd.

7 SRv6

This topic introduces IPv6 Segment Routing (SRv6) and basic principles of SRv6 VPN, and provides an example for configuring an SRv6 VPN.

7.1 Feature Change History

7.2 Introduction

7.3 Basic Principles

7.4 IS-IS for SRv6

7.5 SRv6 VPN

7.6 Example for Configuring of SRv6 VPN
This section provides an example for configuring an SRv6 tunnel to carry IPv4 VPN services.

7.7 Example for Configuring SRv6 EVPN

7.1 Feature Change History

Product Version	Change Description
V100R023C10	Added: **7.7 Example for Configuring SRv6 EVPN**
V100R019C1X and later versions	First version to support this feature

7.2 Introduction

Definition

Segment Routing IPv6 (SRv6) is a protocol designed to forward IPv6 data packets on a network based on source routes. IPv6 forwarding plane-based SRv6 enables

 Copyright © Huawei Technologies Co., Ltd.

the ingress to add a segment routing header (SRH) into IPv6 packets. An explicit IPv6 address stack is pushed into the SRH. Transit nodes continue to update IPv6 destination IP addresses and offset the address stack to implement per-hop forwarding.

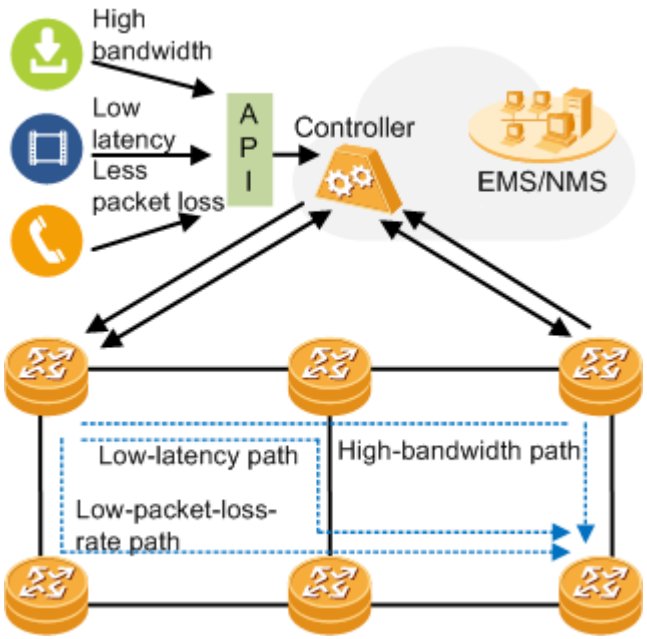
Purpose

Nowadays, different types of network services keep emerging, posing various requirements on the network. For example, real-time UC&C applications require low-latency and low-jitter paths, while big data applications require high-bandwidth paths with low packet loss rate. Traditional network adaptation methods are overstretched in meeting these new requirements, and make network deployment and O&M more and more difficult.

Simplified, low-latency, and SDN/NFV-based networks are the future network development trend. To meet the development requirements of networks, users require simpler VPN implementation using IPv6 addresses.

Segment Routing IPv6 (SRv6) came about in this context. **Figure 7-1** illustrates the network architecture defined by different services. SRv6 uses existing IPv6 forwarding technologies and extends IPv6 packet headers to implement label forwarding-like processing. Using SRv6, some IPv6 addresses are defined as instantiated segment IDs (SIDs). Each SID has its own explicit functions. SIDs are operated to flexibly plan paths and implement simplified VPNs.

Figure 7-1 Service-driven network



Benefits

SRv6 offers the following benefits to users:

- Streamlines network configurations to more easier to implement VPNs.

  SRv6 does not use MPLS techniques and is fully compatible with existing IPv6 networks. Nodes merely support IPv6 forwarding, not MPLS forwarding.

 Copyright © Huawei Technologies Co., Ltd.

Transit nodes can be incapable of SRv6 and forward IPv6 packets carrying the SRH over routes.

- Facilitates traffic optimization on IPv6 forwarding paths.

  SIDs with various service types are used to flexible plan explicit paths on the ingress to adjust service traffic.

- Better smooth evolution to SDN network.

  Segment Routing supports both traditional networks and SDN networks. It is compatible with existing equipment and ensures smooth evolution of existing networks to SDN networks instead of subverting existing networks.

7.3 Basic Principles

SRH

An IPv6 packet consists of a standard IPv6 header, extended headers (0...n), and payload. To implement Segment Routing IPv6 (SRv6) based on the IPv6 forwarding plane, an IPv6 extension header, called segment routing header (SRH), is added. An SRH specifies an explicit path and stores IPv6 segment list information.

The ingress adds an SRH to an IPv6 packet, and each transit node forwards the packet based on path information carried in the SRH. **Figure 7-2** shows the SRH header format.

Figure 7-2 SRH format

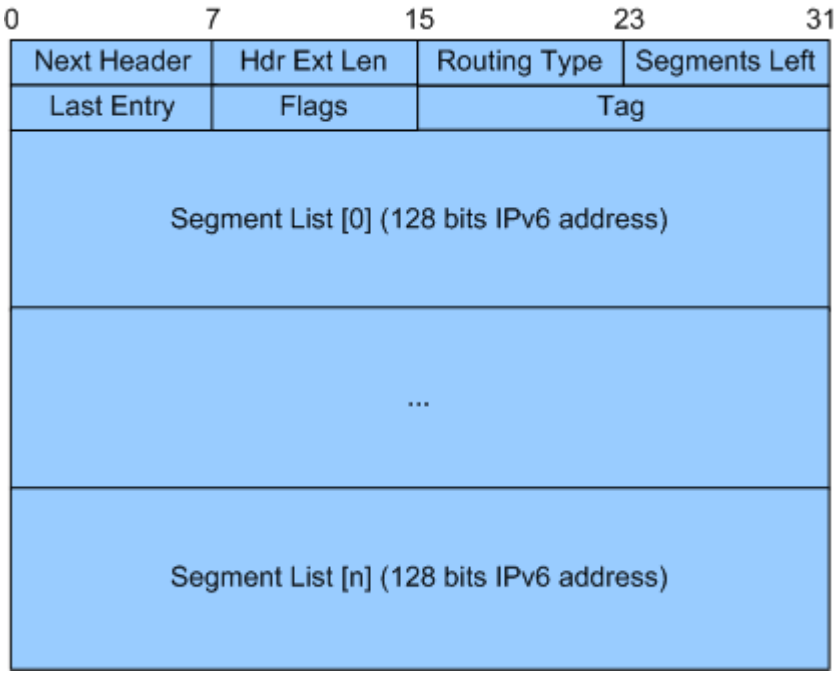


Table 7-1 describes the fields carried in an SRH.

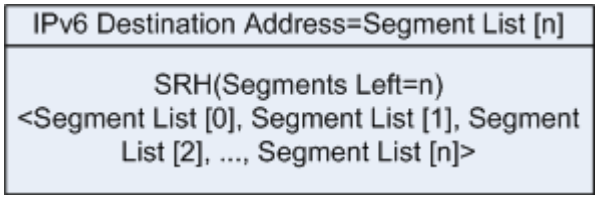
 Copyright © Huawei Technologies Co., Ltd.

Table 7-1 Meanings of fields in an SRH

Field Name	Length	Description
Next Header	8 bits	Type of header immediately following the SRH.
Hdr Ext Len	8 bits	SRH header length. It covers the length from Segment List [0] to Segment List [n].
Routing Type	8 bits	Route header type. Value 4 identifies an SRH type.
Segments Left	8 bits	Number of transit nodes between the existing node and the egress.
Last Entry	8 bits	Index of the last segment in a segment list.
Flags	8 bits	Identifiers in a data packet.
Tag	16 bits	A tag indicating a packet as part of a class or group of packets.
Segment List[n]	128xn bits	Label segment list. A segment list is numbered from the last segment of a path. The Segment List is in the format of an IPv6 address.

Figure 7-3 shows the abstract SRH format.

Figure 7-3 Abstract SRH format



The meaning of each field is as follows:

- IPv6 Destination Address: IPv6 destination address in a packet. It is also called IPv6 DA. The IPv6 DA is a fixed value in an ordinary IPv6 packet. In SRv6, an IPv6 DA only identifies the next hop of an existing packet and is changeable.
- <Segment List [0], Segment List [1], Segment List [2],..., Segment List [n]>: SRv6 packet segment list. Similar to an MPLS label stack in SR MPLS, it is generated on the ingress. The Segment List is encoded starting from the last segment of the path. For example, the first element of the segment list (Segment List [0]) contains the last segment of the path, the second element contains the penultimate segment of the path and so on.

In SRv6, each time a packet passes through an SRv6 node, the Segments Left (SL) field value decreases by 1, and the IPv6 DA changes. Both the Segments Left and Segments List fields determine IPv6 DA information.

 Copyright © Huawei Technologies Co., Ltd.

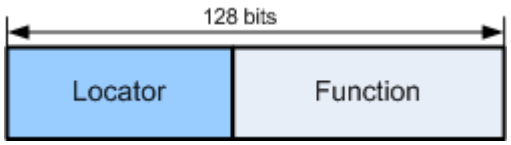
- If the SL value is n, the IPv6 DA value is equal to the Segments List [n] value.
- ...
- If the SL value is 1, the IPv6 DA value is equal to the Segments List [1] value.
- If the SL value is 0, the IPv6 DA value is equal to the Segments List [0] value.

SRv6 Segment

An SRv6 Segment is a 128-bit value, which is also called an SRv6 SID or SID. A SID is an instantiated IPv6 address that is assigned a unique function. An SRv6 SID identifies a node, a link, an L2VPN or L3VPN, or a service. The SRv6 SID can define any network function.

An SRv6 SID is a network instruction that consists of the locator and function. The locator is responsible for routing and must be unique within an SR domain. The function identifies any function on a device, such as a forwarding action or a service. The SRv6 SID structure facilitates network programming. **Figure 7-4** shows the SRv6 SID structure.

Figure 7-4 SRv6 SID



SRv6 SIDs are of various types. Each type of SRv6 SID indicates a specific function. End SID and End.X SID are the most common SRv6 SIDs. An End.DT4 SID stands for an IPv4 VPN. **Table 7-2** lists SRv6 segment types.

Table 7-2 SRv6 segment types

Label	Generation Mode	Function	Example
End SID	Manually configured.	Endpoint SID, which is used to identify the prefix of a destination address on a network. The End SID is similar to the Prefix Segment in SR MPLS. An IGP floods the End SID to the other NEs. The End SID is visible globally and takes effect globally.	**Figure 7-5**
End.X SID	Manually configured.	Endpoint SID of a Layer 3 cross-connection, which is used to identify a link on a network. The End.X SID is similar to Adjacency Segment in SR MPLS. An IGP floods the End.X SID to the other NEs. The End.X SID is visible globally and takes effect locally.	**Figure 7-6**

 Copyright © Huawei Technologies Co., Ltd.

Label	Generation Mode	Function	Example
End.DT4 SID	Automatically generated path name.	An End.DT4 SID stands for a PE endpoint SID that identifies an IPv4 VPN instance on a network. The forwarding behavior mapped to an End.DT4 SID is to decapsulate packets and searches the routing table of an IPv4 VPN instance for an entry to forward the packets. The End.DT4 SID is an equivalence to an IPv4 VPN label used in L3VPN scenarios.	Figure 7-7

Figure 7-5 End SID

Figure 7-6 End.X SID

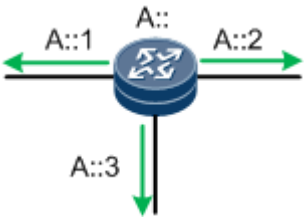
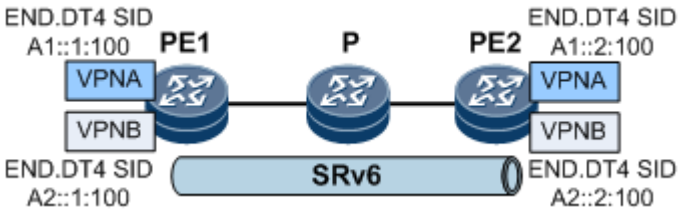


Figure 7-7 End.DT4 SID



An SRv6-capable node maintains a local SID table. This table contains all SRv6 segment information generated by the local node. Based on the table, the local node generates an SRv6 forwarding table (FIB). The local SID table provides the following functions:

- Defines a locally generated SID, for example, End.X SID.
- Specifies bindings to the SID.
- Stores related parameters.

Data Forwarding

SRv6 data is forwarded based on End SIDs and End.X SIDs. After receiving SRv6 packets, a node searches the local SID table based on IPv6 destination addresses (IPv6 DAs) and checks whether End or End.X SIDs are used.

 Copyright © Huawei Technologies Co., Ltd.

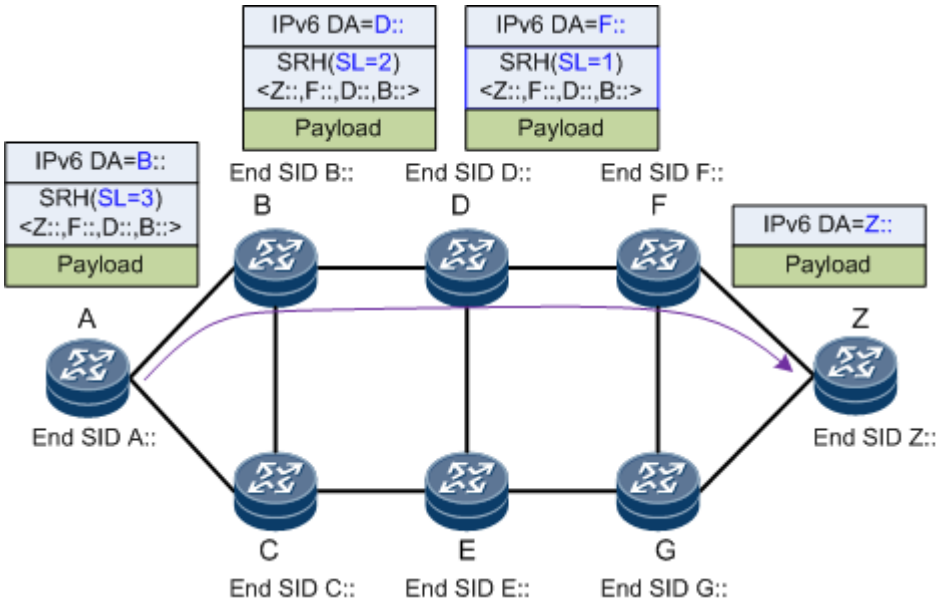
- If End SIDs are used, the node searches the IPv6 FIB table, finds a matching outbound interface and a next hop, and forwards the packets.
- If End.X DISs are used, the node forwards the packets through an outbound interface to a next hop specified in an End.X SID.

End SID-based data forwarding

Figure 7-8 demonstrates the process of forwarding data based on End SIDs:

1. Node A pushes an SRH into a data packet. The path information is <Z::, F::, D::, B::>, and the IPv6 DA is B::.
2. Each time the packet passes through a node, for example, nodes B and D, a node searches the local SID table based on the IPv6 DA, checks the End type, searches the IPv6 FIB table, finds a matching outgoing interface and next hop, reduces the SL value by one, and changes the IPv6 DA at a time.
3. After the packet arrives at node F, node F searches the local SID table based on the IPv6 DA, checks that the End type is used, continues to query the IPv6 FIB table, and finds the outgoing interface and next-hop IP address. In addition, node F reduces the SL value to 0 and changes the IPv6 DA to Z::. The path information <Z::, F::, D::, B::> becomes meaningless, and therefore, node F uses the PSP to remove SRH path information and forward the packet to node Z.

Figure 7-8 End SID-based data forwarding



End.X SID-based data forwarding

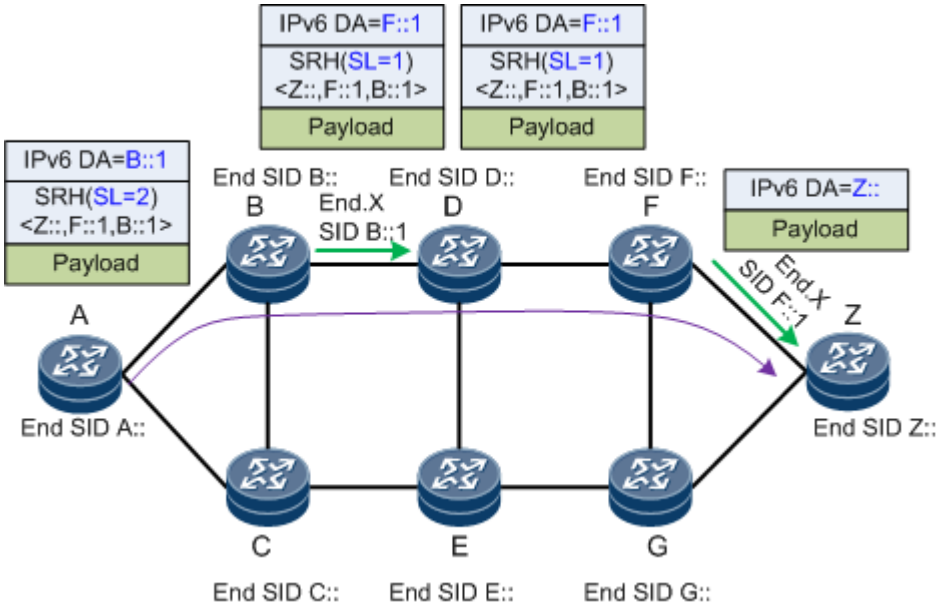
Figure 7-9 demonstrates the process of forwarding data based on End.X SIDs:

1. Node A pushes an SRH into a data packet. The path information is <Z::, F::1, B::1>, and the IPv6 DA is B::1.
2. After the packet arrives at node B, node B searches the local SID table based on the IPv6 DA, checks that the End.X type is used, forwards the packet through the outgoing interface to the next hop, reduces the SL value by one, and changes the IPv6 DA to F::1.

 Copyright © Huawei Technologies Co., Ltd.

3. After the packet arrives at node D, node D searches the local SID table based on the IPv6 DA of F::1, finds no match, continues to use the IPv6 DA to query the FIB table, and finds a matching entry to forward the packet to node F.
4. After the packet arrives at node F, node F searches the local SID table based on the IPv6 DA, checks that the End.X type is used, forwards the packet through the outgoing interface to the next hop, reduces the SL value to 0, and changes the IPv6 DA to Z::. The path information <Z::, F::1, B::1> becomes meaningless, and therefore, node F uses the PSP to remove SRH path information and forward the packet to node Z.

Figure 7-9 End.X SID-based data forwarding



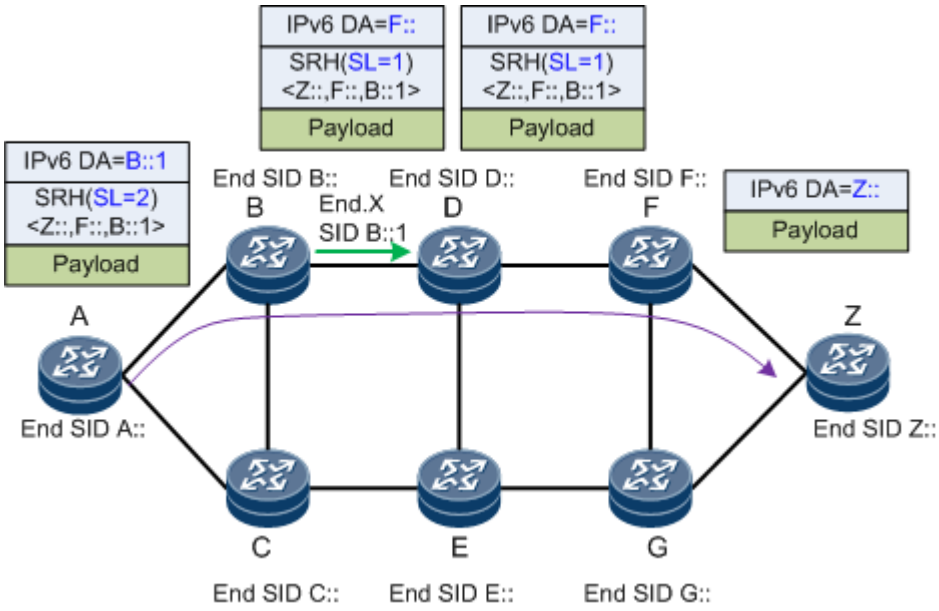
Data forwarding based on End SIDs and End.X SIDs

Figure 7-10 demonstrates the process of forwarding data based on End SIDs and End.X SIDs:

1. Node A pushes an SRH into a data packet. The path information is <Z::, F::, B::1>, and the IPv6 DA is B::1.
2. After the packet arrives at node B, node B searches the local SID table based on the IPv6 DA, checks that the End.X type is used, forwards the packet through the outgoing interface to the next hop, reduces the SL value by one, and changes the IPv6 DA to F::.
3. After the packet arrives at node D, node D searches the local SID table based on the IPv6 DA of F::, finds no match, continues to use the IPv6 DA to query the FIB table, and finds a matching entry to forward the packet to node F.
4. After the packet arrives at node F, node F searches the local SID table based on the IPv6 DA, checks that the End type is used, forwards the packet through the outgoing interface to the next hop, reduces the SL value to 0, and changes the IPv6 DA to Z::. The path information <Z::, F::, B::1> becomes meaningless, and therefore, node F uses the PSP to remove SRH path information and forward the packet to node Z.

 Copyright © Huawei Technologies Co., Ltd.

Figure 7-10 Data forwarding based on End SIDs and End.X SIDs



SRH Information Popping

SRv6 supports the penultimate segment POP of the SRH (PSP) function, which is similar to Penultimate Hop Popping (PHP) in MPLS. PSP is enabled to allow the penultimate segment node to remove the SRH that becomes meaningless before a packet is sent to the egress, which reduces the burden on the egress. If PSP is disabled, the SRH is removed on the egress.

7.4 IS-IS for SRv6

Segment routing uses an IGP to advertise topology information, prefix information, a segment routing global block (SRGB), and label information. To complete the preceding functions, the IGP extends some TLVs of protocol packets. **Table 7-3** describes TLVs of the IS-IS SRv6 extension.

Table 7-3 TLVs of the IS-IS SRv6 extension

Type	Name	Function	Position
TLV	SRv6 SID TLV	Used to advertise SRv6 SIDs.	IS-IS packets.
Sub-TLV	SRv6 Capabilities sub-TLV	Used to advertise SRv6 capabilities.	IS-IS TLV of type 242.
	P2P SRv6 X-SID sub-TLV	Used to advertise SRv6 SIDs on a P2P network.	IS-IS TLV of types 22, 23, 141, 222, 223, and so on.
	LAN SRv6 X-SID sub-TLV	Used to advertise SRv6 SIDs on a LAN.	IS-IS TLV of types 22, 23, 141, 222, 223, and so on.

 Copyright © Huawei Technologies Co., Ltd.

Type	Name	Function	Position
Sub-sub-TLV	SRv6-function Descriptor	Used to advertise SRv6 SID function types.	SRv6 SID TLV, P2P SRv6 X-SID sub-TLV, and LAN SRv6 X-SID sub-TLV. In the SRv6 SID TLV, this field functions as a sub-TLV. In the P2P SRv6 X-SID sub-TLV and LAN SRv6 X-SID sub-TLV, this field functions as a sub-sub-TLV.

SRv6-Function Descriptor

The SRv6-function Descriptor defines the function type bound to each SRv6 SID. Figure 7-11 shows the format of the SRv6-function Descriptor.

Figure 7-11 SRv6-function Descriptor Sub-sub-TLV

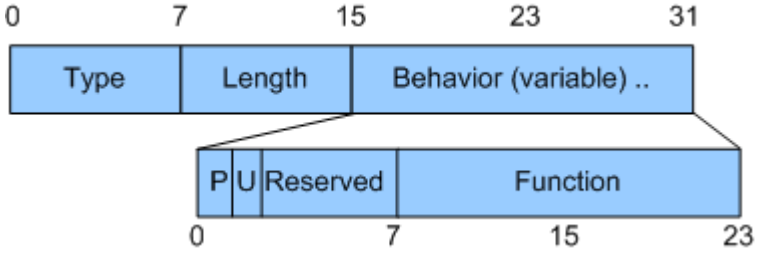


Table 7-4 Meanings of fields in the SRv6-function Descriptor Sub-sub-TLV

Field Name	Length	Description
Type	8 bits	Type.
Length	8 bits	Length. The value is three times the number of functions.

 Copyright © Huawei Technologies Co., Ltd.

Field Name	Length	Description
Behavior	Variable length	Function flag: • P: PSP flag • U: USP flag • Reserved: Reserved bits • Function: function type: – 0: Reserved – 1: End – 2: End.DX6 – 3: End.DT6 – 4: End.X

7.5 SRv6 VPN

SRv6 VPN transmits VPN data along SRv6 tunnels. **Table 7-5** describes the comparison between the SRv6 VPN and BGP/MPLS IPv6 VPN.

Table 7-5 Comparison between the SRv6 VPN and BGP/MPLS IPv6 VPN

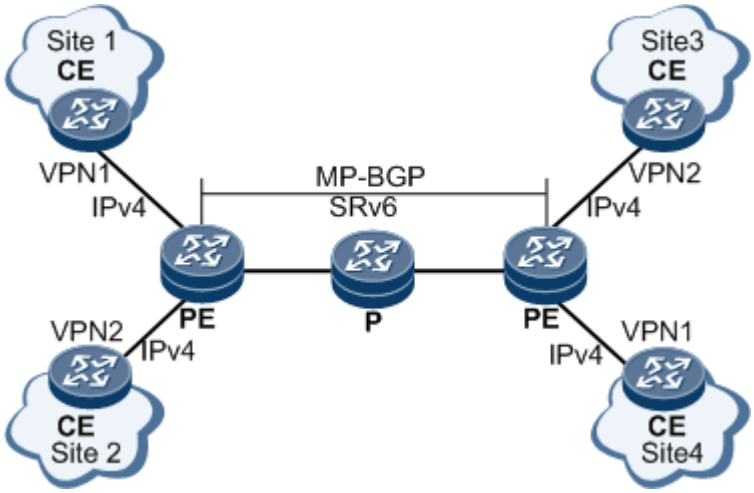
Item	SRv6 VPN	BGP/MPLS IPv6 VPN
VPN service type	IPv4 VPN over SRv6 tunnel	IPv6 VPN over MPLS tunnel
VPNv4 route identifier and crossing	RD: identifies a specified VPN address space. RT: the local import RT must be the same as the peer export RT.	RD: identifies a specified VPN address space. RT: the local import RT must be the same as the peer export RT.
Route transfer	The IPv6 peer relationship is enabled in the BGP VPNv4 address family to transfer IPv4 route information.	The BGP VPNv4 address family is used to transfer IPv6 route information.
Public network route interworking	BGP + IS-IS	BGP + IGP or static route
VPN label	VPN labels do not exist. SRv6 VPN SIDs are used instead.	BGP assigns VPN labels. VPN labels are inner labels carried in BGP/MPLS IPv6 VPN public-network packets and used to identify VPN instances.

 Copyright © Huawei Technologies Co., Ltd.

Item	SRv6 VPN	BGP/MPLS IPv6 VPN
MPLS label	MPLS labels do not exist.	MPLS LDP or TE assigns labels. MPLS labels are outer labels carried in BGP/MPLS IPv6 VPN public-network packets and used to identify tunnels.
Private network table lookup	The egress removes SRv6 VPN SIDs, identifiers VPN instances based on SRv6 VPN SIDs, and searches the local SID table of each VPN instance.	The egress removes MPLS labels and VPN labels, identifies VPN instances based on VPN labels, searches VPN routing tables, and forwards packets over IP.

Figure 7-12 shows the typical SRv6 VPN networking.

Figure 7-12 Typical SRv6 VPN networking



SRv6 VPN characteristics are as follows:

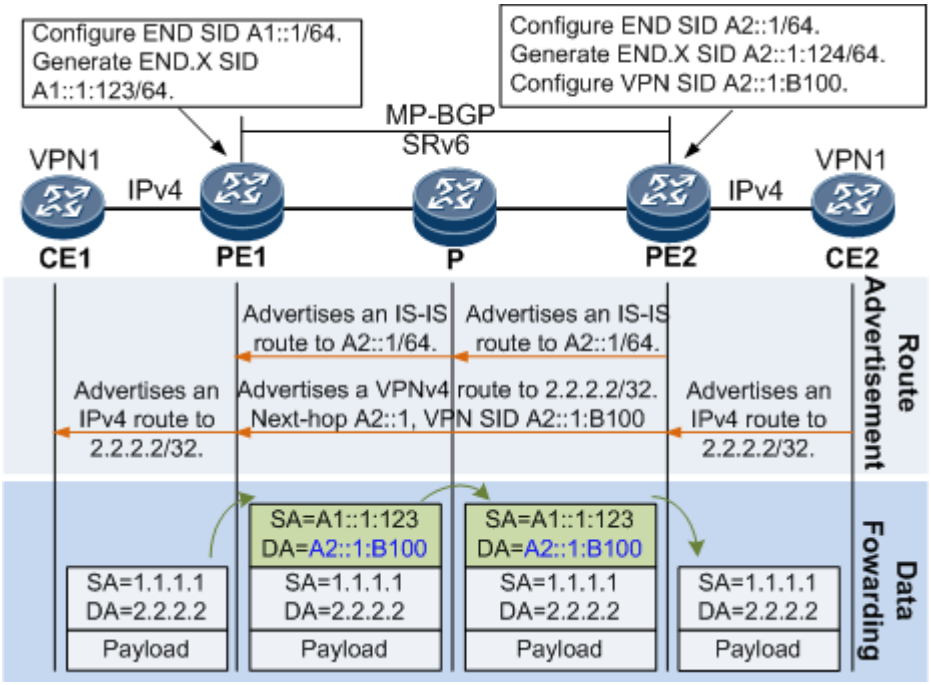
- Transmits packets using extended BGP.
- Encapsulates and transmits private network data packets over SRv6 LSPs or TE tunnels serving as public network tunnels.
- Allows a device that may play PE, P, and CE roles to play only a single role at a time.

IPv4 VPN over SRv6 allows SRv6 on public networks to carry IPv4 VPN data.

Figure 7-13 describes the process of advertising the routes and forwarding data in IPv4 VPN over SRv6.

 Copyright © Huawei Technologies Co., Ltd.

Figure 7-13 Process of advertising the routes and forwarding data in IPv4 VPN over SRv6 LSP



In the route advertisement phase:

1. An END.SID is set on PE2 and advertised to PE1 using an IGP.
2. PE1 runs an IGP and learns the route (A2:1::/64) to the network mapped to the SRv6 SID.
3. PE2 automatically generates an END.DT4 SID (A2:1::B100) within the END.SID range for the VPN instance and a local SID table.
4. After receiving the VPNv4 route advertised by CE2, PE2 converts it to a BGP VPNv4 route and advertises the route to PE1 through a BGP peer relationship. The route carries the SRv6 VPN SID, which is END.DT4 SID (A2:1::B100) of the VPN instance.
5. Upon receipt of the VPNv4 route, PE1 injects the route to the VRF routing table, converts it to a common IPv4 route, and advertises it to CE1.

In the data forwarding phase:

1. CE1 sends a common IPv4 packet to PE1.
2. After receiving the private network packet on an interface bound to a VPN instance, PE1 searches a VRF routing table for an IPv4 prefix matching the destination and finds an associated SRv6 VPN SID and next hop information. PE1 encapsulates the SRv6 VPN SID (A2:1::B100) as a destination address into the IPv6 packet.
3. PE1 finds the route to A2:1/64 based on the longest match rule and forwards the packet to the P over the shortest path.
4. The P finds the route to A2:1/64 based on the longest match rule and forwards the packet to PE2 over the shortest path.
5. PE2 searches the local SID table for a behavior matching A2:1::B100, removes the IPv6 packet header, matches the packet against a VPN instance based on

 Copyright © Huawei Technologies Co., Ltd.

the END.DT4 SID, and searches for a matching entry in the VRF routing table for forwarding. The packet is restored to a common IPv4 packet.

7.6 Example for Configuring of SRv6 VPN

This section provides an example for configuring an SRv6 tunnel to carry IPv4 VPN services.

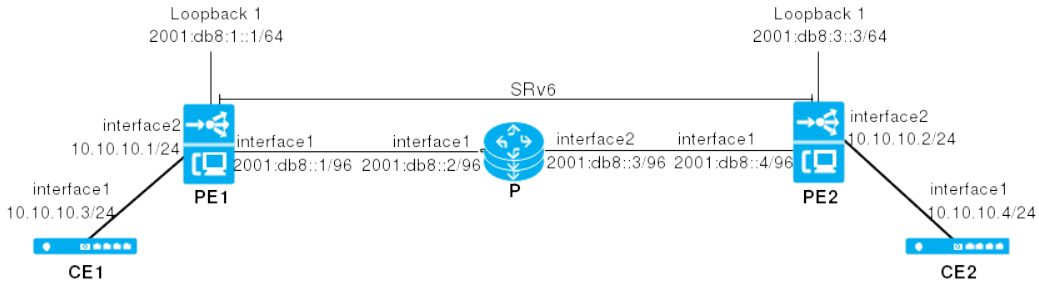
Service Requirements

IPv4 VPN over SRv6 indicates that an IPv6 tunnel on the public network is used to carry private network data of an IPv4 VPN.

As shown in **Figure1**,

- PE1, P and PE2 belong to the same AS and run IS-IS to implement IPv6 network connectivity.
- PE1, P, and PE2 are Level-1 devices in area 1.

Figure 7-14 SRv6 VPN networking diagram



Configuration Flowchart

1. On each device, enable IS-IS, configure a level, and specify a network entity.
2. On PE1 and PE2, configure a VPN instance.
3. Between PE1 and the PE2, set up an IBGP peer relationship.
4. Between PE1 and PE2, configure an SRv6 tunnel for forwarding private network data.

Data preparation

To complete the configuration, you need the following data:

- IPv6 address and VLAN for each interface of PE1, P and the PE2.
- IPv4 address and VLAN for each interface of CE1 and CE2.
- Levels of PE1, P and PE2.
- AS number of PE1, P and the PE2.
- Name and RD of a VPN instance on PE1 and PE2.

Table 7-6 provides the SRv6 VPN configuration data plan.

 Copyright © Huawei Technologies Co., Ltd.

Table 7-6 SRv6 VPN configuration data plan

Item	Data
IS-IS parameters of PE1	IS-IS process ID: 1 Level: level-1 Network entity title: 10.0000.0000.0001.00
	IS-IS interface: Interface ID: 0/9/0 VLAN ID: 20 VLAN type: standard VLAN IPv6 address: 2001:db8::1/96
VPN parameters of PE1	Instance name: vrf100 RD: 100:1 Export VPN target: 1:1 Import VPN target: 1:1
	VPN interface: Upstream interface: 0/9/1 VLAN ID: 200 VLAN type: smart VLAN IPv4 address: 10.10.10.1/24
BGP parameter of PE1	BGP AS number: 100

Example

1. Configure the IS-IS function.

   #Configure PE1.

```
huawei(config)#vlan 20 standard
huawei(config)#port vlan 20 0/9 0
huawei(config)#interface vlanif 20
huawei(config-if-vlanif20)#ipv6 enable
huawei(config-if-vlanif20)#ipv6 address 2001:db8::1 96
huawei(config)#interface loopback1
huawei(config-if-loopback1)#ipv6 enable
huawei(config-if-loopback1)#ipv6 address 2001:db8:1::1 64
huawei(config-if-loopback1)#quit
huawei(config)#isis 1
huawei(config-isis-1)#ipv6 enable
huawei(config-isis-1)#is-level level-1
huawei(config-isis-1)#network-entity 10.0000.0000.0001.00
huawei(config-isis-1)#quit
huawei(config)#interface vlanif 20
huawei(config-if-vlanif20)#isis ipv6 enable 1
huawei(config-if-vlanif20)#quit
huawei(config)#interface loopback1
huawei(config-if-loopback1)#isis ipv6 enable 1
huawei(config-if-loopback1)#quit
huawei(config)#isis 1
huawei(config-isis-1)#ipv6 enable
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-isis-1)#is-level level-1
huawei(config-isis-1)#network-entity 10.0000.0000.0001.00
huawei(config-isis-1)#quit
huawei(config)#interface vlanif 20
huawei(config-if-vlanif20)#isis ipv6 enable 1
huawei(config-if-vlanif20)#quit
huawei(config)#interface loopback1
huawei(config-if-loopback1)#isis ipv6 enable 1
huawei(config-if-loopback1)#quit
```

#Configure PE2. Configure the IS-IS function for PE2 in the same way as PE1.

#Configure P. Configure IS-IS for router P with reference to router configuration descriptions.

2. On a PE, configure a VPN instance for which the IPv4 address family is enabled, and connect a CE to a PE.

   #Configure PE1.

```
huawei(config)#ip vpn-instance vrf100
huawei(config-vpn-instance-vrf100)#ipv4-family
huawei(config-vpn-instance-vrf100-af-ipv4)#route-distinguisher 100:1
huawei(config-vpn-instance-vrf100-af-ipv4)#vpn-target 1:1 export-extcommunity
huawei(config-vpn-instance-vrf100-af-ipv4)#vpn-target 1:1 import-extcommunity
huawei(config-vpn-instance-vrf100-af-ipv4)#quit
huawei(config)#vlan 200 smart
huawei(config)#port vlan 200 0/9 1
huawei(config)#interface vlanif 200
huawei(config-if-vlanif200)#ip binding vpn-instance vrf100
huawei(config-if-vlanif200)#ip address 10.10.10.1 24
huawei(config-if-vlanif200)#quit
```

   #Configure PE2. Configure a Layer 3 VPN for PE2 in the same way as PE1.

   #Configure CE1. Configure the IP address of interface 1 for CE1.

```
huawei(config)#vlan 200 smart
huawei(config)#interface vlanif 200
huawei(config-if-vlanif200)#ip address 10.10.10.3 24
huawei(config-if-vlanif200)#quit
```

   #Configure CE2. Configure the IP address of interface 1 for CE2 in the same way as CE1.

   Each PE can ping the connected CEs.

3. Between PE1 and the RR, set up an IBGP peer relationship.

   #Configure PE1.

```
huawei(config)#bgp 100
huawei(config-bgp)#peer 2001:db8:3::3 as-number 100
huawei(config-bgp)#peer 2001:db8:3::3 connect-interface loopback 1
huawei(config-bgp)#ipv4-family vpnv4
huawei(config-bgp-af-vpnv4)#peer 2001:db8:3::3 enable
huawei(config-bgp-af-vpnv4)#peer 2001:db8:3::3 prefix-sid
huawei(config-bgp-af-vpnv4)#quit
huawei(config-bgp)#quit
```

   #Configure PE2. Configure PE2 in the same way as PE1.

   After the configuration is complete, run the display bgp vpnv4 all peer command on the PE. The command output shows that the EBGP peer relationship has been established between the PE and RR and the status of the EBGP peer relationship is established.

4. Between PEs, configure an SRv6 tunnel for forwarding private network data.

   #Configure PE1.

```
huawei(config)#segment-routing ipv6
huawei(config-segment-routing-ipv6)#encapsulation source-address 2001:db8:11::11
huawei(config-segment-routing-ipv6)#locator test1 ipv6-prefix 1:: 64 static 32
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-segment-routing-ipv6-locator)#quit
huawei(config-segment-routing-ipv6)#quit
huawei(config)#bgp 100
huawei(config-bgp)#ipv4-family vpn-instance vrf100
huawei(config-bgp-vrf100)#segment-routing ipv6 best-effort
huawei(config-bgp-vrf100)#segment-routing ipv6 locator test1
huawei(config-bgp-vrf100)#quit
huawei(config-bgp)#quit
```

#Configure PE2. Configure PE2 in the same way as PE1.

You can run the **display segment-routing ipv6 locator verbose** command to query the Locator information of the IPv6 Segment Routing. And you can run the **display segment-routing ipv6 local-sid** command to query the Local SID information of the IPv6 Segment Routing.

5. Check the configuration result.

   CEs in the same VPN can ping each other.

7.7 Example for Configuring SRv6 EVPN

This section provides an example for configuring a Segment Routing over IPv6 (SRv6) tunnel to carry Ethernet VPN (EVPN) services.

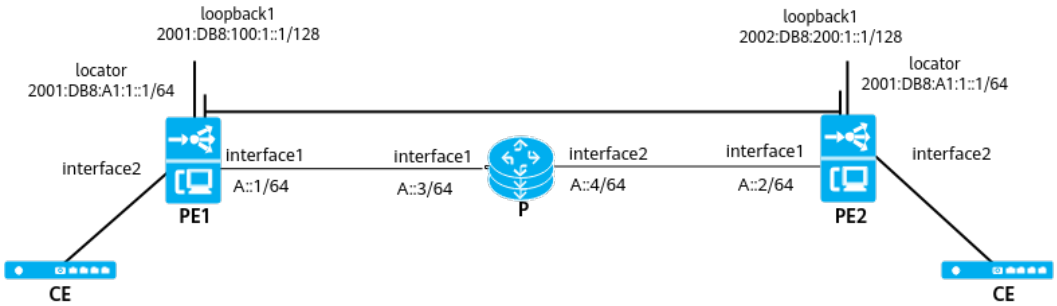
Service Requirement

The EVPN virtual private LAN service (VPLS) uses the EVPN Ethernet local area network (E-LAN) model to carry multipoint-to-multipoint (MP2MP) VPLSs. EVPN VPLS over SRv6 Best Effort (BE) builds on this technology and uses SRv6 BE paths over public networks to carry EVPN E-LAN services. Implementing EVPN VPLS over SRv6 BE involves establishing SRv6 BE paths, advertising EVPN routes, and forwarding data.

As shown in **Figure 7-15**, PE1 and PE2 must communicate through an IPv6 public network and SRv6 BE is deployed on the public IPv6 network to carry EVPN E-LAN services.

- PE1, the P, and PE2 are in the same autonomous system (AS) and run the Intermediate System to Intermediate System (IS-IS) or Open Shortest Path First (OSPF) protocol to implement IPv6 network connectivity.
- In addition, a bidirectional SRv6 BE path is deployed between PE1 and PE2 to carry EVPN E-LAN services.

Figure 7-15 SRv6 EVPN networking diagram



Configuration Procedure

Configure SRv6 EVPN according to the following procedure:

1. Enable IS-IS, configure an IS-IS level, and specify a network entity title (NET) on PE1, the P, and PE2.
2. Configure an EVPN instance working in bridge domain (BD) mode on each PE and bind the corresponding BD to an access-side sub-interface.
3. Establish a BGP EVPN peer relationship between the PEs.
4. Configure SRv6 BE on the PEs.

Data Preparation

To complete the configuration, prepare the following data:

- IPv6 address of each interface on PE1, the P, and PE2 and VLAN planning
- Route distinguisher (RD) for EVPN instance evpn1
- BD ID: 10
- Locator name on PE1: a; dynamically generated opcodes
- Length of the field for locator a: 8 bits (*args*)

Example

1. Enable IS-IS, configure an IS-IS level, and specify an NET on PE1, the P, and PE2.

   # Configure PE1.

   ```
   huawei(config)#vlan 20 standard
   huawei(config)#port vlan 20 0/9 0
   huawei(config)#service-port vlan 20 gpon 0/6/0 ont 0 gemport 1 multi-service user-vlan 20 tag-transform translate
   huawei(config)#vlan 10 standard
   huawei(config)#port vlan 10 0/9 0
   huawei(config)#interface vlanif 10
   huawei(config-if-vlanif10)#ipv6 enable
   huawei(config-if-vlanif10)#ipv6 address A::1/64
   huawei(config-if-vlanif10)#ip address 10.10.10.1 255.255.255.0
   huawei(config)#interface loopback1
   huawei(config-if-loopback1)#ipv6 enable
   huawei(config-if-loopback1)#ipv6 address 2001:DB8:100:1::1/128
   huawei(config-if-loopback1)#quit
   huawei(config)#isis 1
   huawei(config-isis-1)#is-level level-1
   huawei(config-isis-1)#cost-style wide
   huawei(config-isis-1)#network-entity 10.0000.0000.0001.00
   huawei(config-isis-1)#quit
   huawei(config)#interface vlanif 10
   huawei(config-if-vlanif10)#isis ipv6 enable 1
   huawei(config-if-vlanif10)#quit
   huawei(config)#interface loopback1
   huawei(config-if-loopback1)#isis enable 1
   huawei(config-if-loopback1)#isis ipv6 enable 1
   huawei(config-if-loopback1)#quit
   ```

   # Configure PE2. Configure IS-IS on PE2 in the same way as on PE1.

   # Configure the P. Refer to the router configuration for configuring router P.

   After the configuration is complete, run the **display isis peer** command to display the IS-IS neighbor information; run the **display isis route** command to display the routing table information.

 Copyright © Huawei Technologies Co., Ltd.

2. Configure SRv6 BE between the PEs.

   # Configure PE1.

   ```
   huawei(config)#segment-routing ipv6
   huawei(config-segment-routing-ipv6)#encapsulation source-address 2001:DB8:100:1::1
   huawei(config-segment-routing-ipv6)#locator a ipv6-prefix 2001:DB8:A1:1::1 64 args 8
   huawei(config-segment-routing-ipv6-locator)#quit
   huawei(config-segment-routing-ipv6)#quit
   huawei(config)#isis 1
   huawei(config-isis-1)#ipv6 enable topology ipv6
   huawei(config-isis-1)#segment-routing ipv6 locator a
   huawei(config-isis-1)#quit
   huawei(config-evpn-instance-evpn1)#quit
   ```

   # Configure PE2. Configure PE2 in the same way as PE1.

3. Configure an EVPN instance working in BD mode on each PE, and bind the instance to the corresponding BD.

   # Configure PE1.

   ```
   huawei(config)#evpn source-address 1.1.1.1
   huawei(config)#evpn vpn-instance evpn1 bd-mode
   huawei(config-evpn-instance-evpn1)#route-distinguisher 100:1
   huawei(config-evpn-instance-evpn1)#segment-routing ipv6 best-effort
   huawei(config-evpn-instance-evpn1)#segment-routing ipv6 locator a
   huawei(config-evpn-instance-evpn1)#vpn-target 1:1 export-extcommunity
   huawei(config-evpn-instance-evpn1)#vpn-target 1:1 import-extcommunity
   huawei(config-evpn-instance-evpn1)#quit
   huawei(config)#bridge-domain 10
   huawei(config-bd-10)#evpn binding vpn-instance evpn1
   huawei(config-bd-10)#evpn binding vlan 20
   huawei(config-bd-10)#quit
   ```

   # Configure PE2. Configure PE2 in the same way as PE1.

4. Configure a BGP EVPN peer relationship between the PEs.

   # Configure PE1.

   ```
   huawei(config)#bgp 100
   huawei(config-bgp)#private-4-byte-as enable
   huawei(config-bgp)#peer 2002:DB8:200:1::1 as-number 100
   huawei(config-bgp)#peer 2002:DB8:200:1::1 connect-interface LoopBack1
   huawei(config-bgp)#ipv4-family unicast
   huawei(config-bgp-af-ipv4)#undo synchronization
   huawei(config-bgp-af-ipv4)#quit
   huawei(config-bgp)#l2vpn-family evpn
   huawei(config-bgp-af-evpn)#policy vpn-target
   huawei(config-bgp-af-evpn)#peer 2002:DB8:200:1::1 enable
   huawei(config-bgp-af-evpn)#peer 2002:DB8:200:1::1 reflect-client
   huawei(config-bgp-af-evpn)#peer 2002:DB8:200:1::1 advertise encap-type srv6
   huawei(config-bgp-af-evpn)#quit
   huawei(config-bgp)#quit
   ```

   # Configure PE2. Configure PE2 in the same way as PE1.

   After the configuration is complete, run the **display bgp evpn peer** command on the PEs to check whether a BGP EVPN peer relationship has been established between them. The command output shows that the BGP EVPN peer relationship has been established between the PEs and is in the **Established** state.

5. Verify the configurations.

   Run the **display segment-routing ipv6 local-sid** { **end-dt2u** | **end-dt2m** } **forwarding** command on each PE to check information about the local SID table of the SRv6 BE tunnel.

   Run the **display bgp evpn all routing-table** command on each PE. The command output shows the EVPN route sent from the peer device.

 Copyright © Huawei Technologies Co., Ltd.

8 VXLAN

Virtual eXtensible Local Area Network (VXLAN) is a technology for overlaying virtualized Layer 2 networks over Layer 3 networks.

8.1 VXLAN Overview
VXLAN is an important feature of network cloudification. It allows a virtual network to provide access services to a large number of tenants. In addition, tenants are able to plan their own virtual networks, not limited by the physical network IP addresses or broadcast domains. This greatly simplifies network management.

8.2 Basic VXLAN Concepts
Virtual eXtensible Local Area Network (VXLAN) is a Network Virtualization Overlays (NVO3) technology. VXLAN encapsulates data packets sent from VMs into UDP packets using the IP and MAC addresses of the physical devices as the outer headers, and then transmits the UDP packets over IP networks. Upon receiving the packets, the destination tunnel endpoint decapsulates them and sends them to the destination VM or physical server.

8.3 VXLAN Implementation Principles
VXLAN encapsulates packets in MAC-in-UDP mode so that Layer 2 packets can be transmitted on the Layer 3 networks.

8.4 VXLAN Deployment
A VXLAN is deployed on a layer-3 network to implement communication between virtual layer-2 devices.

8.5 Configuring VXLAN
This section describes how to configure VXLAN on the CLI.

8.6 Configuring EVPN-VXLAN

8.1 VXLAN Overview

VXLAN is an important feature of network cloudification. It allows a virtual network to provide access services to a large number of tenants. In addition, tenants are able to plan their own virtual networks, not limited by the physical network IP addresses or broadcast domains. This greatly simplifies network management.

 Copyright © Huawei Technologies Co., Ltd.

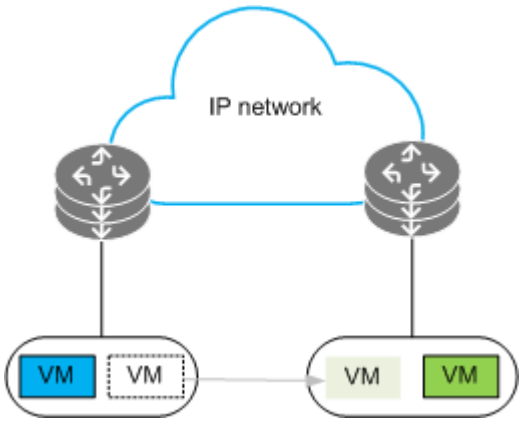
The rapid development of virtualization technologies such as network cloudification faces the following challenges:

- A single physical device is often virtualized into multiple virtual machines for networking. With the increase of application modules, the number of supported VLANs is also increasing. The 802.1Q standard supports a maximum of 4094 VLANs, which cannot meet the requirement of identifying a large number of user groups in a virtual network. This severely limits the number of virtual machines.
- A physical network may be shared by multiple tenants in cloud services, and the tenants may have varying network requirements and overlapping IP and MAC addresses, which cannot be addressed by traditional VLAN technologies.
- After a VM is started, it may need to be migrated to a new server due to resource issues on the original server. For example, the CPU usage is too high, or memory resources are inadequate. To ensure that services are not interrupted during VM migration, the IP address and MAC address of the VM must remain unchanged before and after the migration. This requires that the service network is a Layer 2 network.

The VXLAN technology effectively addresses the preceding problems.

- A 24-bit VXLAN network ID (VNI) is used to identify a network. A maximum of 16777215 (16M) VXLAN segments can be isolated, a lot more than 4096 virtual network identifiers of a VLAN and meeting the requirements of a large number of users.
- The MAC-in-UDP packet encapsulation allows for overlaying virtual Layer 2 networks over Layer 3 networks, decoupling physical networks from virtual networks. The VXLAN network devices are not aware of the VXLAN communication, and the network configurations are easier. Creating VXLANs allows multiple tenants to have independent networks and communicate with each other without IP address conflicts.
- After a large Layer 2 network is created using VXLAN, devices except edge devices in a VXLAN network do not need to identify the VM's MAC address, which reduces the number of MAC addresses learned and enhances network performance. As shown in **Figure 8-1**, tenants can flexibly migrate their own VMs and plan virtual networks without being limited by the number of IP addresses or size of the broadcast domains on physical networks, significantly simplifying network management.

Figure 8-1 VM migration



 Copyright © Huawei Technologies Co., Ltd.

8.2 Basic VXLAN Concepts

Virtual eXtensible Local Area Network (VXLAN) is a Network Virtualization Overlays (NVO3) technology. VXLAN encapsulates data packets sent from VMs into UDP packets using the IP and MAC addresses of the physical devices as the outer headers, and then transmits the UDP packets over IP networks. Upon receiving the packets, the destination tunnel endpoint decapsulates them and sends them to the destination VM or physical server.

Figure 8-2 shows the VXLAN gateway service model.

Figure 8-2 VXLAN service model

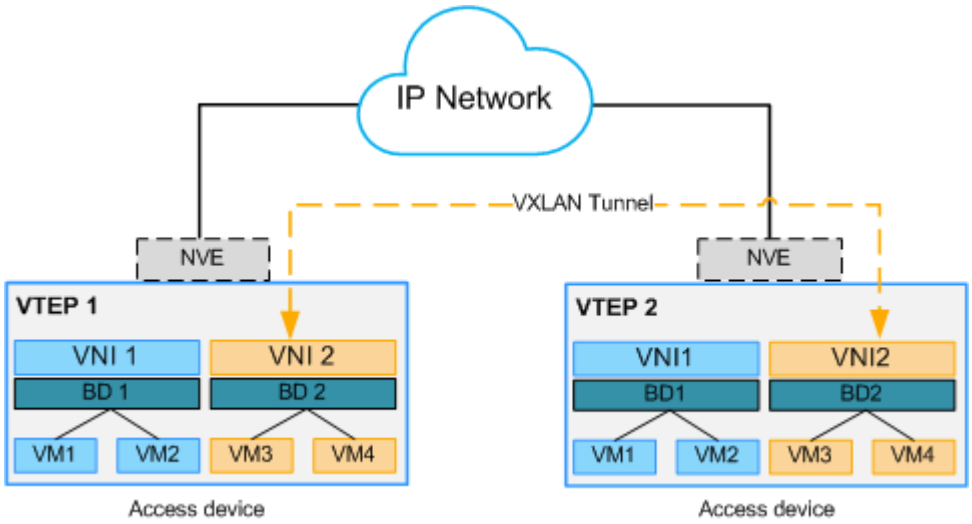


Table 8-1 Basic VXLAN Concepts

Concept	Description
VXLAN tunnel	It is used to forward VXLAN packets. It is identified by the local VTEP address + remote VTEP address.
NVE (Network virtualization edge)	It is the entity for implementing network virtualization. It is used for managing local VTEP addresses, VXLAN tunnels, and ingress replication lists.

 Copyright © Huawei Technologies Co., Ltd.

Concept	Description
VTEP (VXLAN tunnel endpoint)	A device that supports VXLAN is a VTEP. It is encapsulated in NVE for encapsulating and decapsulating VXLAN packets. A VTEP is connected to a physical network and is assigned with a physical network IP address, which is irrelevant to the virtual network. For a VXLAN packet, the source IP address is the VTEP address of the local node and the destination IP address is the VTEP address of the peer node. These two VTEP addresses correspond to a VXLAN tunnel.
BD (Bridge Domain)	It is a Layer 2 broadcast domain for forwarding data packets on a VXLAN. On a VXLAN, VNIs are mapped into the BDs one on one. The BD is the entity for forwarding data packets on the VXLAN network.
VNI (VXLAN network identifier)	It is similar to VLAN ID and is used to identify a VXLAN. VMs in different VXLANs cannot communicate directly. A VNI (24 bits) indicates an ISP and a maximum of 16M ISPs can be supported.

8.3 VXLAN Implementation Principles

VXLAN encapsulates packets in MAC-in-UDP mode so that Layer 2 packets can be transmitted on the Layer 3 networks.

VXLAN packet encapsulation

The MAC-in-UDP packet encapsulation is implemented by adding a VXLAN header to Layer 2 packets and then encapsulating the packets in UDP packets. **Figure 8-3** shows the VXLAN packet encapsulation mode and packet format.

 Copyright © Huawei Technologies Co., Ltd.

Figure 8-3 VXLAN packet format

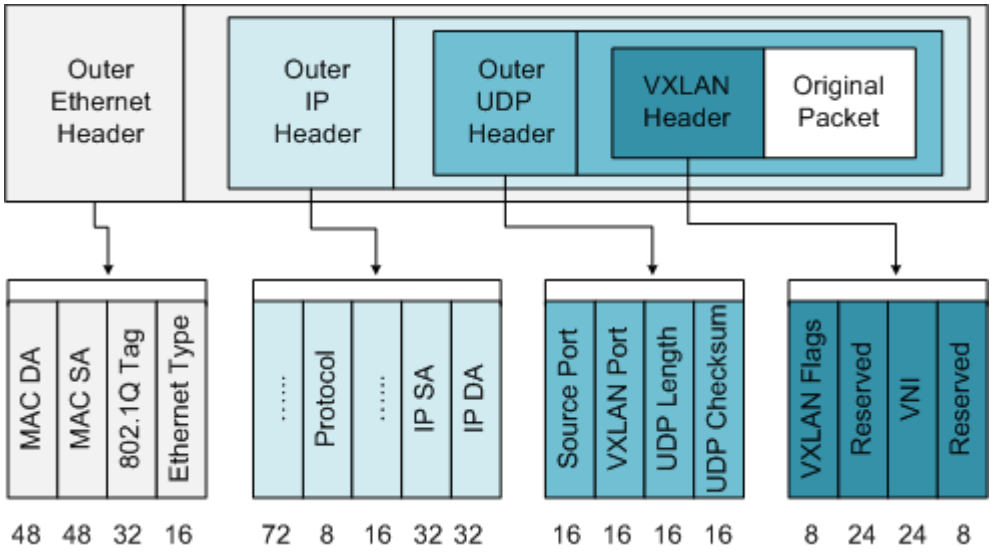


Table 8-2 VXLAN packet description

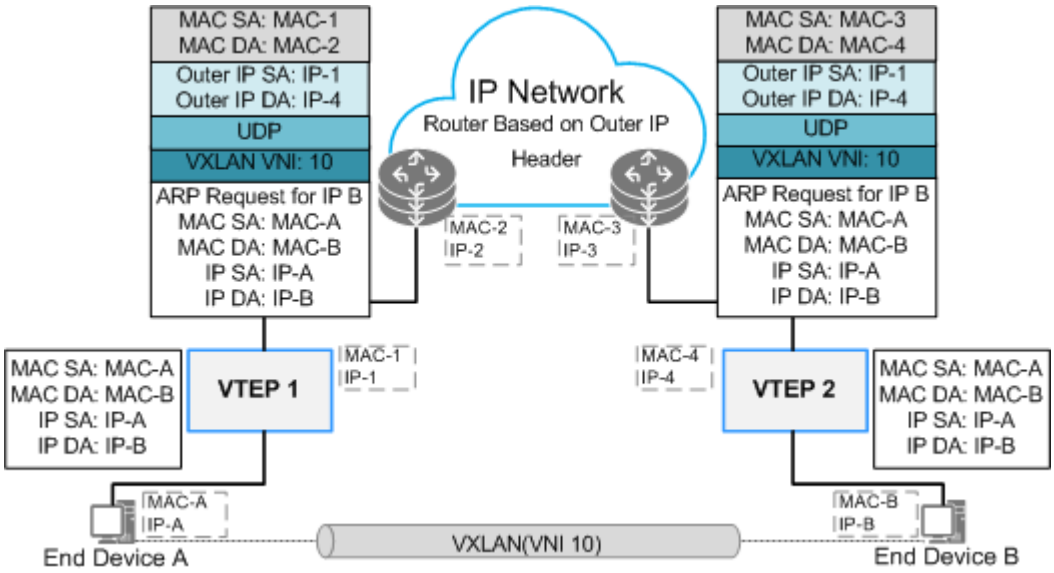
Packet Field	Description
VXLAN Header	VNI: indicates the 24-bit VXLAN identifier, which identifies a VXLAN.
Outer UDP Header	• DestPort: indicates a destination UDP port. • Source Port: indicates the source port.
Outer IP Header	• IP SA: indicates the source IP address, which is the IP address of the VTEP where the sender VM resides. • IP DA: indicates the destination IP address, which is the IP address of the VXLAN Tunnel EndPoint (VTEP) where the destination VM resides.
Outer Ethernet Header	• MAC DA: indicates the MAC address of the next-hop IP address in the routing table of the VTEP where the destination VM resides. • MAC SA: indicates the MAC address of the VTEP where the sender VM resides. • Ethernet Type: indicates the Ethernet packet type.

 Copyright © Huawei Technologies Co., Ltd.

VXLAN packet forwarding process

VXLAN uses VTEPs, such as OLTs, to map user terminals to VXLANs and encapsulate and decapsulate packets to for VXLAN packet transmission. **Figure 8-4** shows the principles of the VXLAN.

Figure 8-4 VXLAN principles



VMs A and B are on the same VXLAN. VM A sends packet M to VM B. The process is described as follows:

1. A sends packet M, which contains the virtual MAC addresses of the sender and receiver.
2. VTEP1 connected to A receives the packet, searches for the VTEP where the sender resides and the VTEP connected to the receiver, adds a VXLAN header, and a UDP header, and then an IP header to the packet, and sends the packet to the destination (VTEP2).
3. The packet is transmitted to the destination (VTEP2) through the physical network.
4. Upon receiving the packet, VTEP2 removes the outer IP header and UDP header of the packet, checks the VNI of the packet and the destination MAC address of the internal data frame, determines that receiver B is connected to the local VTEP, removes the VXLAN header, and sends the packet to B.
5. B receives the packet and the transmission is complete.

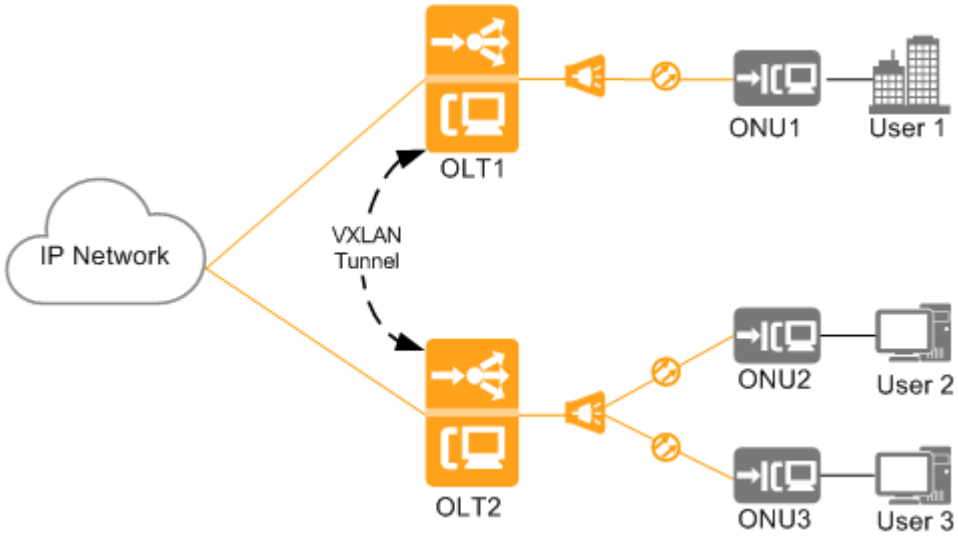
8.4 VXLAN Deployment

A VXLAN is deployed on a layer-3 network to implement communication between virtual layer-2 devices.

The following figure shows the VXLAN networking. OLT1 and OLT2 are located on different layer 2 networks. You can configure VXLANs on the OLTs to implement virtual layer 2 communication between OLT1 and OLT2.

 Copyright © Huawei Technologies Co., Ltd.

Figure 8-5 VXLAN networking



8.5 Configuring VXLAN

This section describes how to configure VXLAN on the CLI.

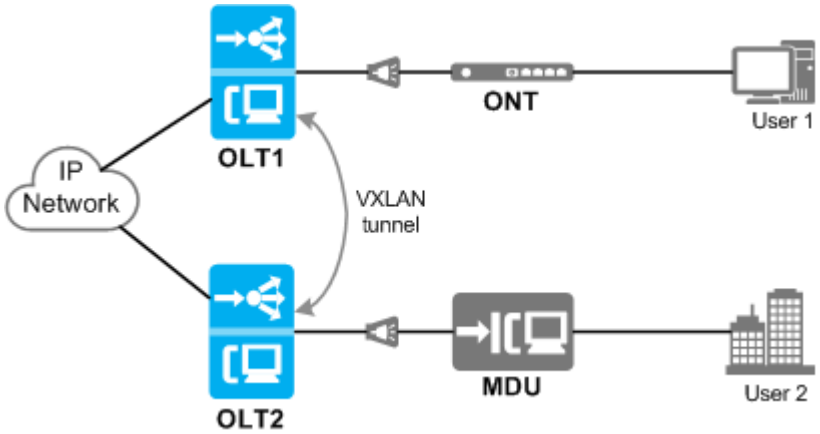
Prerequisites

1. Service configurations (such as VLAN and service flow configurations) have been completed, and user services are normal.
2. The loopback IP address has been configured.

Networking

As shown in **Figure 8-6**, when two OLTs are located on different Layer 2 networks, you can configure a VXLAN for the two OLTs to establish a VXLAN tunnel. In this way, the OLTs can communicate with each other at Layer 2, and Layer 2 packets can be transmitted between User1 and User2.

Figure 8-6 VXLAN tunnel networking



 Copyright © Huawei Technologies Co., Ltd.

The configuration procedures for OLT1 and OLT2 are the same. The following uses OLT1 as an example.

Procedure

Step 1 Run the **interface nve** command to create an NVE interface and enter the NVE view.

Step 2 Run the **source** *ip-address* command to configure an IP address for the source VTEP.

Step 3 Run the **bridge-domain** *bd-id* command to create a bridge domain (BD) and enter the BD view.

Step 4 Run the **vxlan vni** *vni-id* command to create a VXLAN network identifier (VNI).

Step 5 Run the **vni** *vni-id* **vni head-end peer-list** *ip-address* command to configure a VNI ingress replication list and specify the VNI and IP address of the peer VTEP.

Step 6 Run the **bd-binding** *vlan-id* command to bind the VLAN/service port to the BD.

----End

Example

To create vxlan vni 1, associate it with bridge-domain 1, and specify the VNI and the IP address of the peer VTEP of the tunnel as 10.10.11.11, do as follows:

```
huawei(config)#interface nve 1
huawei(config-Nve1)#source 10.10.10.11
huawei(config-Nve1)#vni 1 head-end peer-list 10.10.11.11
huawei(config-Nve1)#quit

huawei(config)#bridge-domain 1
huawei(config-bd-1)#vxlan vni 1
huawei(config-bd-1)#bd-binding vlan 100
huawei(config-bd-1)#quit
```

8.6 Configuring EVPN-VXLAN

EVPN Concepts

Ethernet VPN (EVPN) is a VPN technology based on Layer 2 network interconnection. Similar to BGP/MPLS IP VPN, EVPN uses the reachable information extended through the extended Border Gateway Protocol (BGP), to transfer MAC address learning and advertisement at Layer 2 network of different sites from the data plane to control plane.

EVPN VXLAN Configuration in Single Homing and One Service Flow Working

Figure 8-7 shows the single homing and one service flow working.

 Copyright © Huawei Technologies Co., Ltd.

Figure 8-7 Single homing and one service flow working

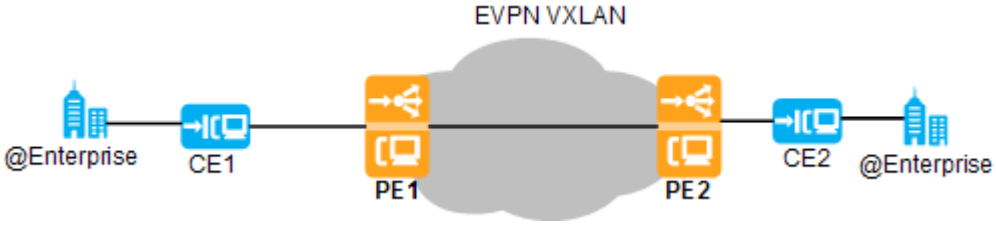


Table 8-3 Device interface parameters

Device	Parameter Configuration
Provider Edge (PE1)	• loopback0 IP: 10.1.1.1 32 • Stand VLAN: 1001, Port 0/9/3 • Smart VLAN: 12, Port 0/9/0 • evpl Route Distinguisher (RD): 101:1 • evpl RT: 200:1
PE2	• loopback0 IP: 10.1.1.2 32 • Stand VLAN: 1001, Port 0/9/3 • Smart VLAN: 12, Port 0/9/0 • Smart VLAN: 32, Port 0/9/1 • evpl RD: 101:1 • evpl RT: 200:1

The following uses PE1 as an example to describe the detailed configuration:

1. In Loopback 0 mode, set the IP address to 10.1.1.1 32.
```
huawei(config)#interface loopback 0
huawei(config-if-loopback0)#ip address 10.1.1.1 32
huawei(config-if-loopback0)#quit
```
2. Add smart VLAN 12 in batches for the Internet access service, and add upstream port 0/9/0 to the VLAN. Add standard-type management VLAN 1001, and add upstream port 0/9/3 to the VLAN.
```
huawei(config)#vlan 12 smart
huawei(config)#vlan 1001 standard
huawei(config)#port vlan 12 0/9 0
huawei(config)#port vlan 1001 0/9 3
```
3. Create a VLANIF interface with IP address 10.1.1.12 24 for VLAN 12.
```
huawei(config)#interface vlanif 12
huawei(config-if-vlanif12)#ip address 10.1.1.12 24
huawei(config-if-vlanif12)#quit
```
4. Run OSPF. Set ID of the area where the interface locates to 0.0.0.0, and specify main IP address of the interface running OSPF to be in network segment 10.1.1.1.
```
huawei(config)#ospf router-id 10.1.1.1
huawei(config-ospf-1)#area 0.0.0.0
huawei(config-ospf-1-area-0.0.0.0)#network 10.1.1.1 0.0.0.0
huawei(config-ospf-1-area-0.0.0.0)#network 10.1.1.13 0.0.0.255
huawei(config-ospf-1-area-0.0.0.0)#quit
huawei(config-ospf-1)#quit
```

 Copyright © Huawei Technologies Co., Ltd.

5. Create EVPN instance evpn1, and set RD to 101:1 and RT to 200:1.
```
huawei(config)#evpn vpn-instance evpn1 vpws
huawei(config-vpws-evpn-instance-evpn1)#route-distinguisher 101:1
huawei(config-vpws-evpn-instance-evpn1)#vpn-target 200:1 both
huawei(config-vpn-instance-evpn1)#quit
```
6. Create an NVE interface, configure IP address for source VTEP, and create a dynamic P2P VXLAN tunnel.
```
huawei(config)#interface nve 1
huawei(config-Nve1)#source 10.1.1.1
huawei(config-Nve1)#vni 1 evpl protocol bgp
huawei(config-Nve1)#quit
```
7. Create an EVPL instance with ID 1 in VXLAN mode, bind it to VNI 1, bind the EVPL instance to EVPN instance, and enable local and remote service ID carried in the EVPL instance.
```
huawei(config)#evpl instance 1 vxlan-mode
huawei(config-evpl-vxlan1)#local-vni 1
huawei(config-evpl-vxlan1)#evpn binding vpn-instance evpn1
huawei(config-evpl-vxlan1)#local-service-id 1 remote-service-id 2
huawei(config-evpl-vxlan1)#evpl-binding vlan 1001
huawei(config-evpl-vxlan1)#quit
```
8. Set the router ID in BGP mode, set the peer AS number of the IPv4 peer 10.1.1.1 to 100, and specify the source interface and source address of BGP packets.
```
huawei(config)#bgp 100
huawei(config-bgp)#router-id 10.1.1.1
huawei(config-bgp)#peer 10.1.1.1 as-number 100
huawei(config-bgp)#peer 10.1.1.1 connect-interface loopback 0
huawei(config-bgp)#peer 10.1.1.1 enable
huawei(config-bgp)#l2vpn-family evpn
huawei(config-bgp-af-evpn)#peer 10.1.1.1 enable
huawei(config-bgp-af-evpn)#quit
huawei(config-bgp)#quit
```

EVPN VXLAN Configuration in Dual Homing and One Service Flow Working

Figure 8-8 shows the dual homing and one service flow working, in which, PE3 is added and CE1 is dual homed to PE1 and PE3 based on single homing and one service flow working.

- Only one of the ACs on PE1 and PE3 can work at a time.
- The ESI of the user-side interface on PE3 must be the same as that on PE1.
- The local and remote service IDs of instance evpl on PE3 must be the same as that on PE1.

Figure 8-8 Dual homing and one service flow working

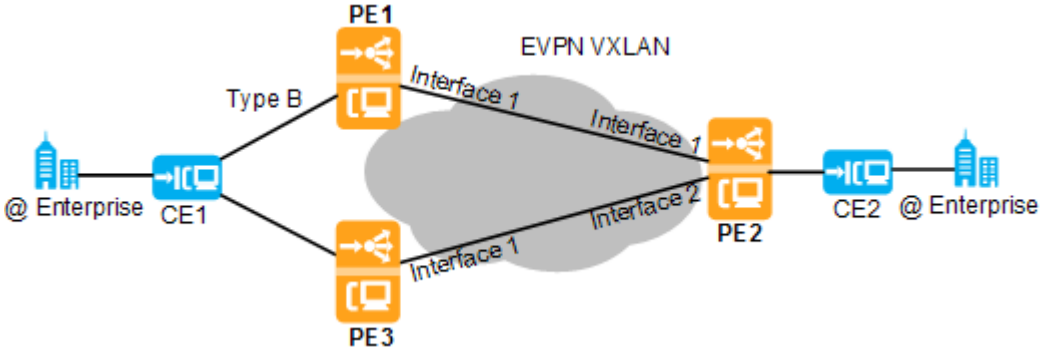


Table 8-4 Device interface parameters

Device	Parameter Configuration
PE1	• loopback0 IP: 10.1.1.1 32 • StandVLAN: 1001, Port 0/9/3 • SmartVLAN: 12, Port 0/9/0 • evpl RD: 101:1 • evpl RT: 200:1
PE2	• loopback0 IP: 10.1.1.2 32 • StandVLAN: 1001, Port 0/9/3 • SmartVLAN: 12, Port 0/9/0 • SmartVLAN: 32, Port 0/9/1 • evpl RD: 101:1 • evpl RT: 200:1
PE3	• loopback0 IP: 10.1.1.3 32 • StandVLAN: 1001, Port 0/9/3 • SmartVLAN: 32, Port 0/9/0 • evpl RD: 101:1 • evpl RT: 200:1

In dual homing and one service flow working, configurations of PE1 and PE3 are the same as those in single homing and one service flow working. The following describes how to configure PE2:

1. In Loopback 0 mode, set the IP address to 10.1.1.2 32.
```
huawei(config)#interface loopback 0
huawei(config-if-loopback0)#ip address 10.1.1.2 32
huawei(config-if-loopback0)#quit
```
2. Add smart VLAN 12 in batches for the Internet access service, and add upstream port 0/9/0 to the VLAN. Add standard-type management VLAN 1001, and add upstream port 0/9/3 to the VLAN.
```
huawei(config)#vlan 12 smart
huawei(config)#vlan 32 smart
huawei(config)#vlan 1001 standard
huawei(config)#port vlan 12 0/9 0
huawei(config)#port vlan 32 0/9 1
huawei(config)#port vlan 1001 0/9 3
```
3. Create a VLANIF interface with IP address 10.1.1.12 24 for VLAN 12.
```
huawei(config)#interface vlanif 12
huawei(config-if-vlanif12)#ip address 10.1.1.14 24
huawei(config-if-vlanif12)#quit
```
4. Run OSPF. Set ID of the area where the interface locates to 0.0.0.0, and specify main IP address of the interface running OSPF to be in network segment 10.1.1.1.
```
huawei(config)#ospf router-id 10.1.1.1
huawei(config-ospf-1)#area 0.0.0.0
huawei(config-ospf-1-area-0.0.0.0)#network 10.1.1.1 0.0.0.0
huawei(config-ospf-1-area-0.0.0.0)#network 10.1.1.13 0.0.0.255
huawei(config-ospf-1-area-0.0.0.0)#network 10.1.1.32 0.0.0.255
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-ospf-1-area-0.0.0.0)#quit
huawei(config-ospf-1)#quit
```

5. Create EVPN instance evpn1, and set RD to 101:1 and RT to 200:1.

```
huawei(config)#evpn vpn-instance evpn1 vpws
huawei(config-vpws-evpn-instance-evpn1)#route-distinguisher 102:1
huawei(config-vpws-evpn-instance-evpn1)#vpn-target 200:1 both
huawei(config-vpws-evpn-instance-evpn1)#remote frr enable
huawei(config-vpn-instance-evpn1)#quit
```

6. Create an NVE interface, configure IP address for source VTEP, and create a dynamic P2P VXLAN tunnel.

```
huawei(config)#interface nve 1
huawei(config-Nve1)#source 10.1.1.1
huawei(config-Nve1)#vni 2 evpl protocol bgp
huawei(config-Nve1)#quit
```

7. Create an EVPL instance with ID 1 in VXLAN mode, bind it to VNI 2, bind the EVPL instance to EVPN instance evpn1, and enable local and remote service ID carried in the EVPL instance.

```
huawei(config)#evpl instance 1 vxlan-mode
huawei(config-evpl-vxlan1)#local-vni 2
huawei(config-evpl-vxlan1)#evpn binding vpn-instance evpn1
huawei(config-evpl-vxlan1)#local-service-id 2 remote-service-id 2
huawei(config-evpl-vxlan1)#evpl-binding vlan 1001
huawei(config-evpl-vxlan1)#quit
```

8. Set the router ID in BGP mode, set the peer AS number of the IPv4 peer 10.1.1.1 to 100, and specify the source interface and source address of BGP packets.

```
huawei(config)#bgp 100
huawei(config-bgp)#router-id 10.1.1.1
huawei(config-bgp)#peer 10.1.1.1 as-number 100
huawei(config-bgp)#peer 10.1.1.1 connect-interface loopback 0
huawei(config-bgp)#peer 10.1.1.1 enable
huawei(config-bgp)#peer 10.1.1.3 as-number 100
huawei(config-bgp)#peer 10.1.1.3 connect-interface loopback 0
huawei(config-bgp)#peer 10.1.1.3 enablehuawei(config-bgp)#l2vpn-family evpn
huawei(config-bgp-af-evpn)#peer 10.1.1.1 enable
huawei(config-bgp-af-evpn)#peer 10.1.1.3 enable
huawei(config-bgp-af-evpn)#quit
huawei(config-bgp)#quit
```

Reference Standard

Reference standards: RFC7432 and RFC4360.

 Copyright © Huawei Technologies Co., Ltd.

9 VPN

Introduce the VPN protocols and features.

9.1 PWE3
PWE3 is an end-to-end Layer 2 service carrying technology and is a type of point-to-point L2VPN technology. PWE3 is proposed by the IETF PWE3 working group as one of the solutions for connecting the traditional communication network with the PSN network.

9.2 Native TDM
In Native TDM, TDM frames are directly encapsulated to GPON GEM frames. This mode features simple encapsulation, small network cost, and guaranteed link quality. TDMoGEM mode and TDMoETH mode are encapsulated to NativeTDM in access network at present, and the MA5800 supports TDMoETH mode.

9.3 VPLS
The Virtual Private LAN Service (VPLS), also called the Transparent LAN Service (TLS) or virtual private switched network service, is a Layer 2 VPN (L2VPN) technology that is based on Multi-Protocol Label Switching (MPLS) and Ethernet technologies.

9.4 VRF
Virtual route forwarding instance (VRF) is a mechanism in which a device works as multiple virtual routing devices. After the Layer 3 interfaces of the device are divided into different VRFs, multiple route forwarding instances can be emulated on the device.

9.5 BGP/MPLS IP VPN

9.6 GRE

9.1 PWE3

PWE3 is an end-to-end Layer 2 service carrying technology and is a type of point-to-point L2VPN technology. PWE3 is proposed by the IETF PWE3 working group as one of the solutions for connecting the traditional communication network with the PSN network.

 Copyright © Huawei Technologies Co., Ltd.

9.1.1 Introduction to PWE3

Definition

Pseudo-wire emulation edge to edge (PWE3) is a type of Layer 2 service carrying technology. It is mainly used to emulate the essential behavior and characteristics of the services such as the ATM, frame relay, Ethernet, low-rate time division multiplexing (TDM) circuit, and synchronous optical network (SONET)/ synchronous digital hierarchy (SDH) as faithfully as possible in a packet switched network (PSN).

PWE3 is implemented on access devices through MPLS and IP technologies. MPLS supports PWE3 by using the LDP or RSVP-TE protocol as signaling.

Purpose

PWE3 can interconnect the traditional network with PSN to share resources and expand the reach of networks. For example, PWE3 can emulate services such as TDM, ATM, and Ethernet, and can implement service interoperation by using existing PSN (IP/MPLS) as the bearer network.

Benefit

PWE3 connects the traditional TDM, ATM, and Ethernet networks with PSN (IP/ MPLS). In this way, PWE3 protects the investment on the traditional TDM, ATM, and Ethernet networks, and also implements the all-IP network architecture.

9.1.2 PWE3 Reference Standards and Protocols

The following lists the reference standards and protocols of this feature.

- RFC 3985: Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture
- RFC 4447: Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)
- RFC 3916: Requirements for Pseudo-Wire Emulation Edge-to-Edge (PWE3)
- RFC 4446: IANA Allocations for Pseudo wire Edge to Edge Emulation (PWE3)
- RFC 4717, Encapsulation Methods for Transport of Asynchronous Transfer Mode (ATM) over MPLS Networks
- RFC 4816, Pseudowire Emulation Edge-to-Edge (PWE3) Asynchronous Transfer Mode (ATM) Transparent Cell Transport Service
- RFC 4448: Encapsulation Methods for Transport of Ethernet over MPLS Networks
- RFC 5085: PW vccv A control Channel for PW
- RFC 4553: Structure-Agnostic Time Division Multiplexing (TDM) over Packet (SAToP)
- RFC 5462: Multi-Protocol Label Switching (MPLS) Label Stack Entry EXP Field Renamed to Traffic Class Field
- RFC 4385: Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN

 Copyright © Huawei Technologies Co., Ltd.

- RFC 6870: Pseudowire Preferential Forwarding Status Bit
- draft-bryant-filsfils-fat-pw-03.txt

9.1.3 PWE3 Principle

9.1.3.1 Basic Principle of PWE3

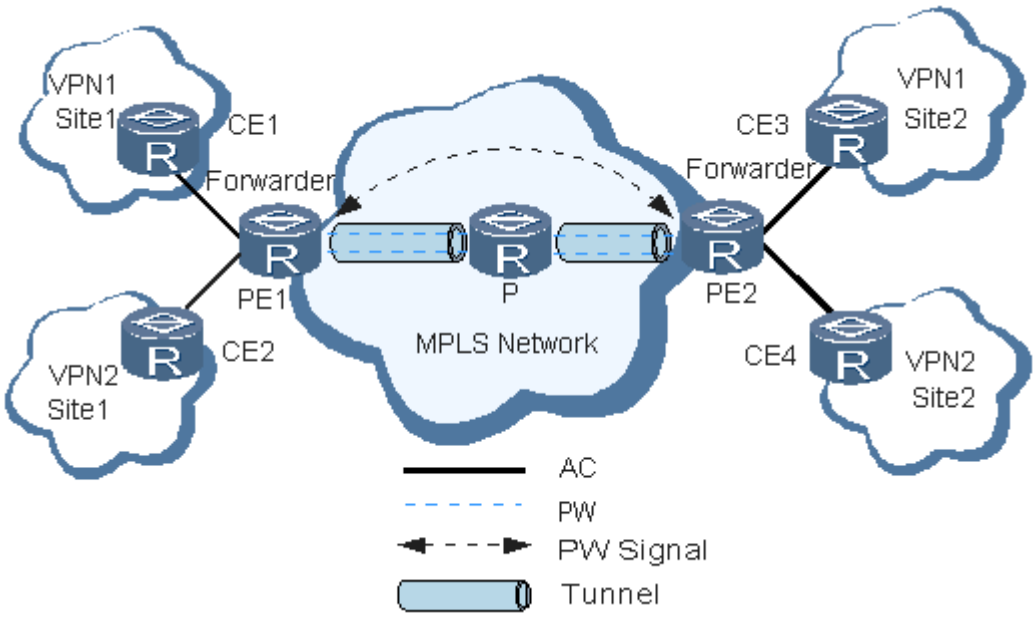
Basic PWE3 Transmission Components

Pseudo wire emulation edge-to-edge (PWE3), which uses LDP and RSVP-TE as the signaling protocols, carries various types of Layer 2 services, such as various types of Layer 2 data packets, from the customer edge (CE), and transparently transmits the Layer 2 data through tunnels (such as MPLS LSP or TE tunnels). As shown in **Figure 9-1**, the basic PWE3 transmission components include the following:

- Attachment circuit (AC): a link between CE and PE. All user packets (including Layer 2 and Layer 3 protocol packets of users) on the AC are transparently forwarded to the peer end.
- Pseudo wire (PW): PW is a kind of virtual connection between two PEs. It is a mechanism for two PEs in the packet-switched network to transfer essential elements of emulation service. A PW conveys VC information by signaling (LDP or RSVP-TE). Since VC is directional, PW is directional too. For the PWE3 system, a PW is like a direct channel between a local AC and a peer AC and is used for transparently transmitting the Layer 2 data of users.
- Forwarder: After a PE receives data frames from an AC, the forwarder selects a PW for forwarding the frames. In fact, the forwarder is a forwarding table of PWE3.
- Tunnel: A tunnel is a direct channel between a local PE and a peer PE and is used for transparently transmitting data between the PEs. Tunnels are used for carrying PWs. A tunnel can carry multiple PWs. Generally, the tunnel refers to an MPLS tunnel.
- PW signaling protocol: A PW signaling protocol is the basis for implementing PWE3 and is used to create and maintain PWs. Current PW signaling protocols are mainly LDP and RSVP-TE.
- Encapsulation: The packets transmitted through the PW use the standard PW encapsulation format and technology. There are multiple PWE3 encapsulation types on a PW. The formats are defined in detail in draft-ietf-pwe3-iana-allocation-x.
- Quality of service (QoS): The priority information at the header of Layer 2 user packets is mapped to the QoS priority for transmitting the packets in the public network. In general, support for MPLS QoS is required.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-1 Basic PWE3 transmission components



Assume that the VPN1 packet stream travels from CE1 to CE3. The basic data flow would be as follows:

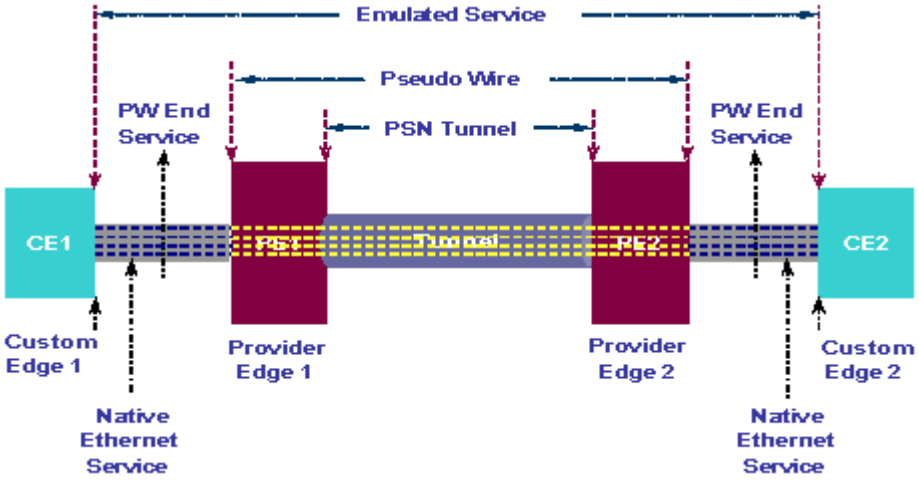
- CE1 transmits a Layer 2 packet to PE1 through an AC.
- After PE1 receives the packet, the forwarder selects a PW for forwarding the packet.
- PE1 generates two MPLS labels according to the PW forwarding entry. The private network label is used for identifying the PW, and the public network label is used for transmitting the packet to PE2 through the tunnel.
- The Layer 2 packet arrives at PE2 through the public network tunnel. The system extracts the private network label (the public network label is extracted by the penultimate P device).
- The forwarder of PE2 selects an AC for forwarding the packet, and then PE2 forwards the packet to CE3.

PWE3 Network Model

Figure 9-2 shows a PWE3 reference model.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-2 PWE3 network model



The channel set up in a PWE3 network is a point-to-point channel. Channels are isolated from each other. Layer 2 user packets are transparently transmitted between PWs. The following provides a detailed description.

- According to the services requirements of the CE, one or more PWs are set up between PE1 and PE2. Multiple PWs can be carried on one PSN tunnel.
- For the PEs, after the PW is set up, the mapping between the user access interface (AC) and virtual link (PW) is determined.
- The PSN device only needs to forward the MPLS packet according to the MPLS label, regardless of the Layer 2 user packet encapsulated inside the MPLS packet.

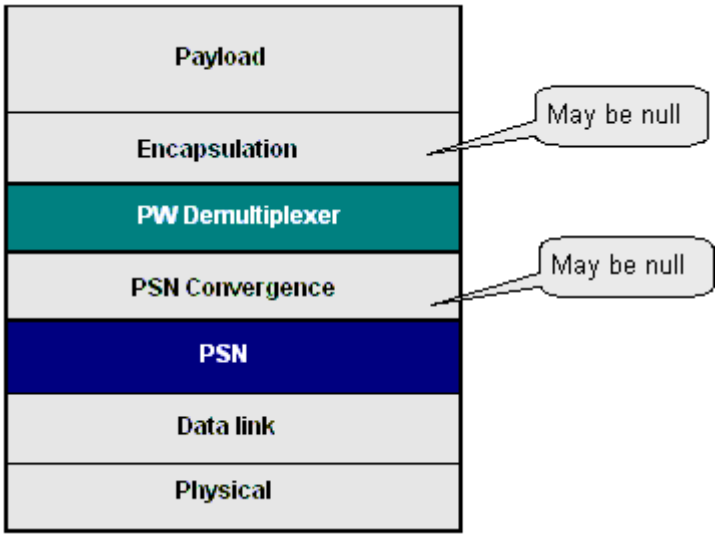
PWE3 Service Model

Figure 9-3 shows a PWE3 service model. According to the PWE3 service model, PWE3 is presented by an outer label (PSN tunnel) and an inner label (PW demultiplexer).

The PSN layer can adopt the MPLS and IP technologies, and the PW demultiplexer layer can adopt the MPLS, UDP, or L2TP technology. Hence, the supported combinations of PWE3 outer labels and inner labels are as follows: MPLS over MPLS, MPLS over IP, UDP over IP, and L2TP over IP.

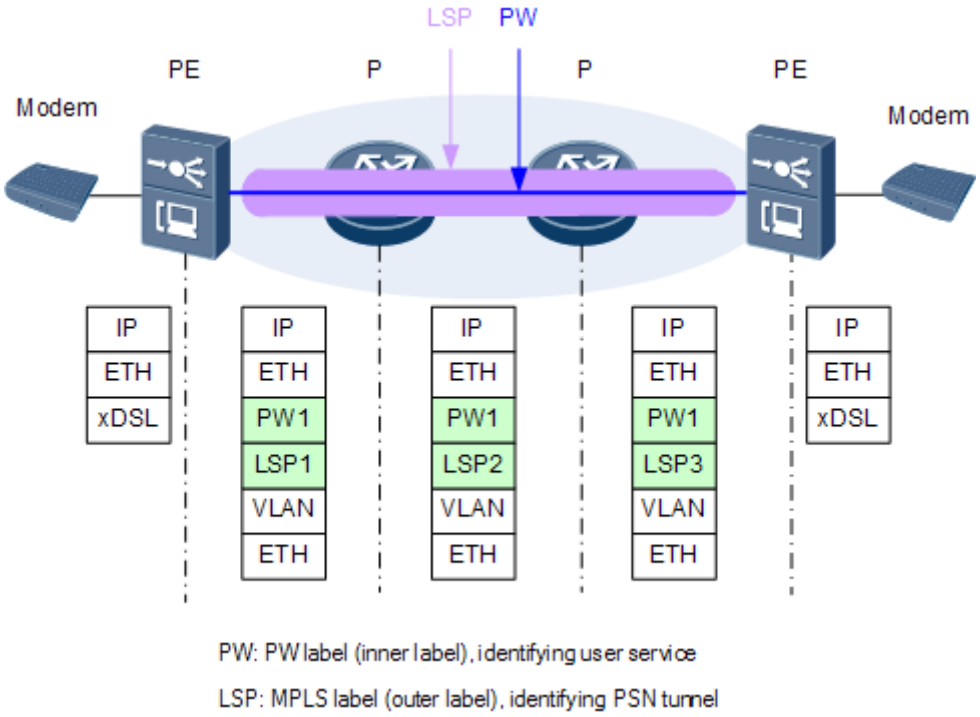
 Copyright © Huawei Technologies Co., Ltd.

Figure 9-3 PWE3 service model



- **Figure 9-4** shows the PWE3 protocol stack in the MPLS over MPLS encapsulation mode.

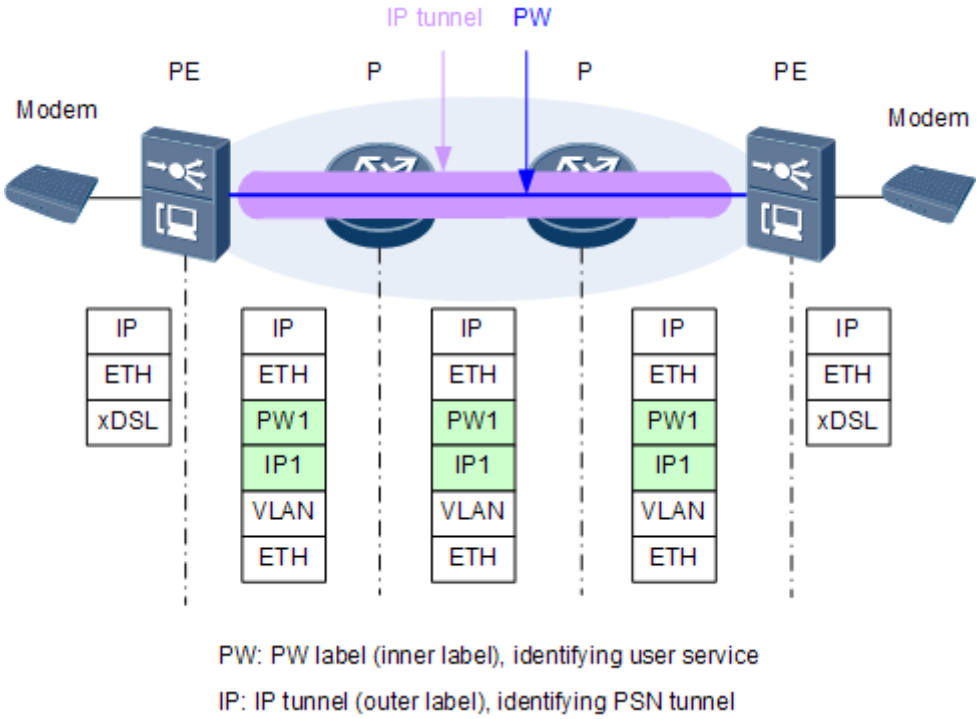
  Figure 9-4 PWE3 protocol stack in the MPLS over MPLS encapsulation mode



- **Figure 9-5** shows the PWE3 protocol stack in the MPLS over IP encapsulation mode.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-5 PWE3 protocol stack in the MPLS over IP encapsulation mode



- **Figure 9-6** shows the PWE3 protocol stack in the UDP over IP encapsulation mode.

Figure 9-6 PWE3 protocol stack in the UDP over IP encapsulation mode

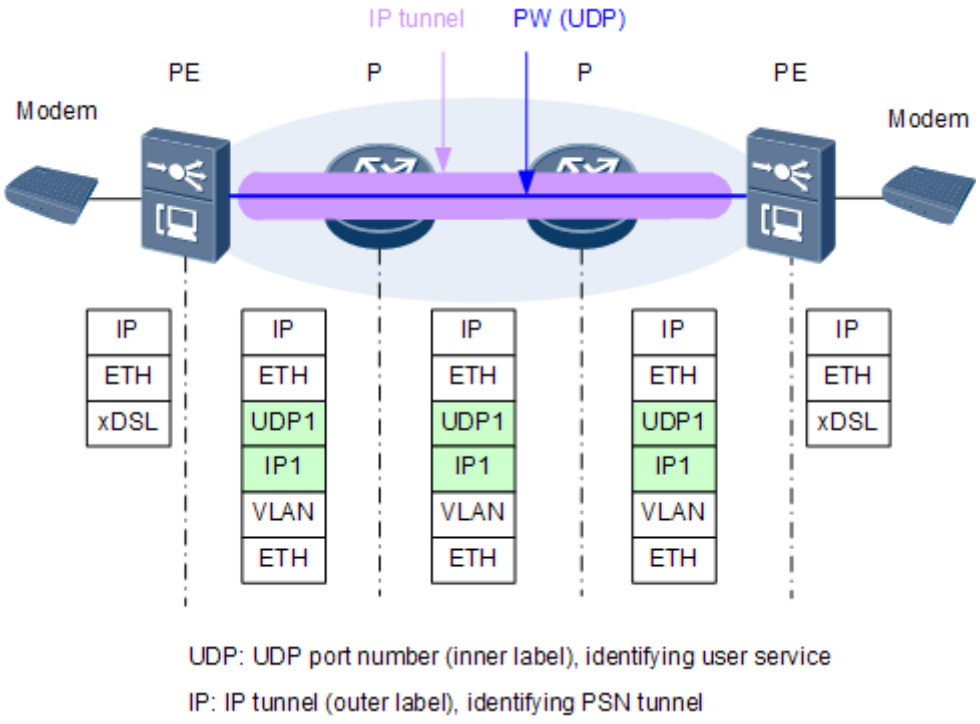
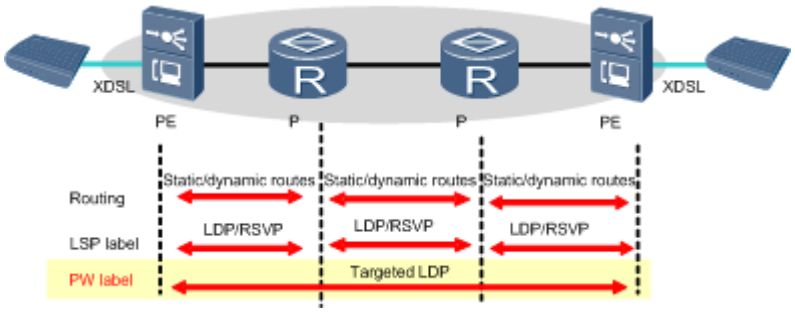


Figure 9-7 illustrates the principle of PW label distribution.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-7 Principle of PW label distribution



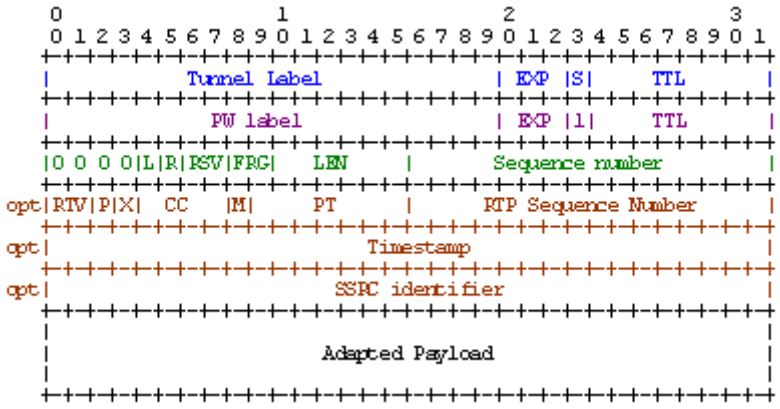
Packet Format

PWE3 has three packet formats: MPLS+PW, IP+PW, and IP+UDP.

- MPLS+PW: In this packet format, the combination of PWE3 outer label and inner label is MPLS over MPLS. It is applicable to MPLS network transmission.

  Figure 9-8 shows the format of an MPLS+PW PWE3 packet.

  Figure 9-8 Format of an MPLS+PW PWE3 packet

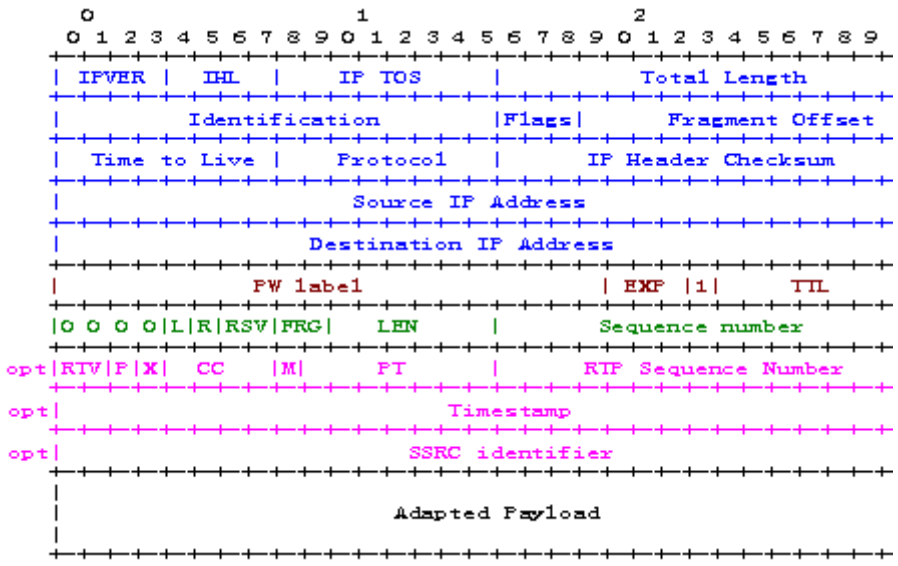


- IP+PW: In this packet format, the combination of PWE3 outer label and inner label is MPLS over IP. It is applicable to MPLS over IP network transmission. Different from MPLS+PW packets, IP+PW packets are forwarded at IP Layer 3 when transmitted over the PSN network.

  Figure 9-9 shows the format of an IP+PW PWE3 packet.

 Copyright © Huawei Technologies Co., Ltd.

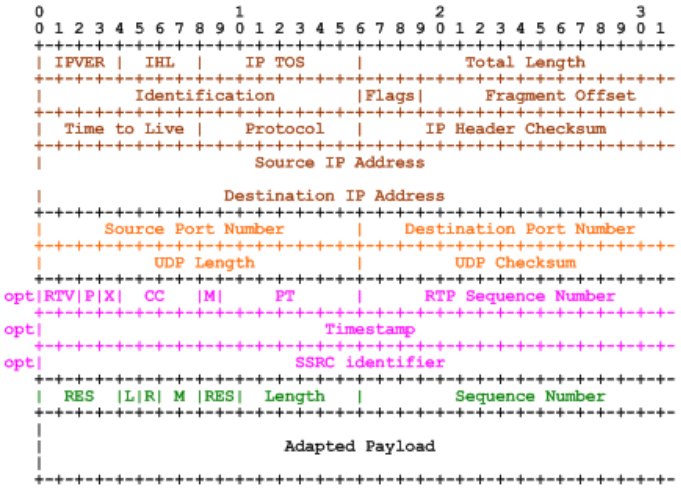
Figure 9-9 Format of an IP+PW PWE3 packet



- IP+UDP: In this packet format, the combination of PWE3 outer label and inner label is UDP over IP. It is applicable to IP network transmission. Different from MPLS+PW packets, IP+UDP packets are forwarded at IP Layer 3 and meanwhile forwarded at Layer 2 by UDP port redirection when transmitted over the PSN network.

  Figure 9-10 shows the format of an IP+UDP PWE3 packet.

Figure 9-10 Format of an IP+UDP PWE3 packet



9.1.3.2 Principle of TDM PWE3

The TDM service is transmitted over the PSN by circuit emulation. There are two packet encapsulation formats.

 Copyright © Huawei Technologies Co., Ltd.

- The structure-aware packet, which is also called CESoPSN packet, is defined in RFC5086 and RFC5687.

  NOTE

  Circuit emulation service over packet switched network (CESoPSN) is a generic term for circuit emulation service and also refers to structure-aware circuit emulation.

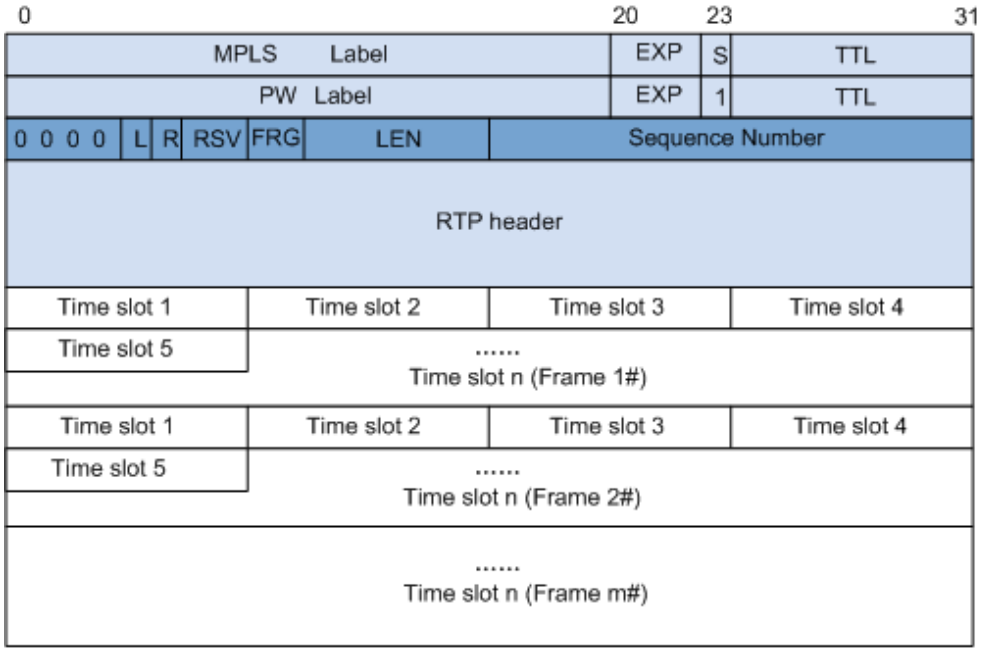
- The structure-agnostic packet, which is also called SAToP packet, is defined in RFC4553.

CESoPSN Packet Format (structure-aware packet format)

The CESoPSN standard provides the channelized TDM service with emulation and transmission functions, and can identify the TDM frame structure and in-frame signaling. Therefore, if the customer needs to provide services based on the timeslot, the CESoPSN packet format can meet this requirement.

Figure 9-11 shows the format of the CESoPSN packet.

Figure 9-11 CESoPSN packet format



- A CESoPSN packet contains a four-byte MPLS header and a four-byte PW header. The length of the CESoPSN control word is 4 bytes, including fields as shown in **Table 9-1**.
- The length of the Real-Time Transport Protocol (RTP) header is 12 bytes, including the version number, padding flag, and time stamp fields. The time stamp field, whose length is 32 bits, is used for clock synchronization. For format of the RTP header, see RFC3550.
- Time slot indicates the TS in the TDM frame. Each TS occupies 8 bits. All TSs comprise the encapsulated TDM data payload, which does not include the CRC bit. The number of encapsulated frames and the number of TSs in each frame can be set by users according to conditions.

 Copyright © Huawei Technologies Co., Ltd.

Table 9-1 Fields of the control word

Field	Description
0000	This field is generally all 0s, with the length of 4 bits. When the virtual circuit connectivity verification (VCCV) is needed to help to monitor the SAToP PW status, these four bits are used to identify the start of the associated channel header (ACH).
L	Indicates whether the TDM data in the packet is valid. Its length is 1 bit. When it is set to 1, it indicates that the TDM data in the packet is invalid; that is, the TDM data in the packet can be neglected to save bandwidth resources.
R	Indicates whether the interconnection function of the local customer edge (CE) is in the packet loss state. When it is set to 0, it indicates that consecutive packets have been received and will no longer be lost.
RSV	Indicates the reserved bit. Its length is 2 bits.
FRG	Indicates the fragmentation status of the packet. Its length is 2 bits. Its values are as follows: • 00: Indicates that the packet encapsulates the entire TDM data. • 01: Indicates that the packet encapsulates the first fragmentation of the TDM data. • 10: Indicates that the packet encapsulates the last fragmentation of the TDM data. • 11: Indicates that the packet encapsulates the intermediate fragmentation of the TDM data.
LEN	Indicates the length of the entire CESoPSN packet (the size of the CESoPSN header and TDM data.) When the length is shorter than 64 bytes, LEN is a specific length value. When the length is equal to or longer than 64 bytes, LEN is 0.
Sequence Number	Indicates the sequence number for transmitting the CESoPSN packet. Its length is 16 bits. Its initial value should be random. It must be incremented by 1 with each CESoPSN data packet sent in the specific PW.

The CESoPSN packet structure has the following characteristics:

- CESoPSN provides emulation and transmission of structure-aware TDM service. That is, CESoPSN can identify the TDM frame structure and in-frame signaling and transmit the frames.

  For example, a structure-aware E1 link consists of 32 timeslots. Except timeslot 0, the other 31 timeslots can each carry a channel of 64 kbit/s voice service. Timeslot 0 is used only for transmitting signaling and frame delimiter.

 Copyright © Huawei Technologies Co., Ltd.

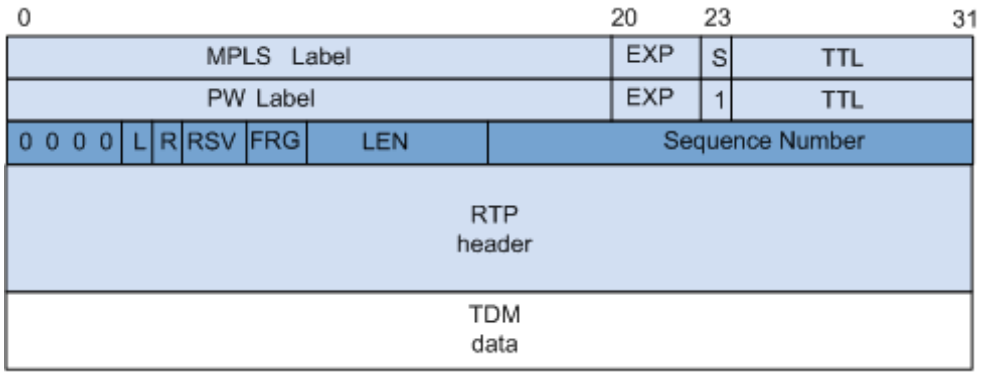
- CESoP can identify the TDM frame structure. Therefore, idle timeslots are not transmitted, and only the data in the timeslots that are useful for the CE devices is retrieved from the E1 service stream and transmitted after being encapsulated into CESoPSN frames.
- CESoPSN can identify and transmit the CAS and CCS signaling of E1 service streams.

SAToP Packet Format (structure-agnostic packet format)

The structure-agnostic TDM over PSN (SAToP) standard provides the non-channelized TDM service with the emulation and transmission functions. The protocol need not be aware of the structure of the TDM packets and transparently transmits the packets. Therefore, if the customer only needs to provide services based on E1/T1, SAToP (unstructured packet format) can meet this requirement.

Figure 9-12 shows the format of the SAToP packet.

Figure 9-12 SAToP packet format



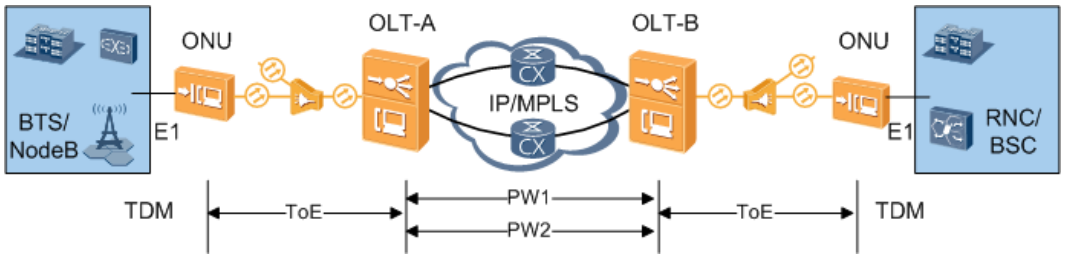
- An SAToP packet also contains a four-byte MPLS header and a four-byte PW header. The length of the SAToP control word is 4 bytes, including fields as shown in Table 9-1.
- The SAToP protocol treats the TDM service as serial data code stream for segmentation, and transmits the service over PWs after encapsulation. SAToP can transmit the synchronization timing information although it is unaware of the structure of the TDM frame.

Application Scenario

As shown in Figure 9-13, OLT-A and OLT-B are symmetrically deployed on both sides of the PSN network. Both OLT-A and OLT-B are connected to ONUs through GPON lines. Enterprise or base transceiver station (BTS) private line services are transmitted to the ONUs through E1 ports. Proprietary TOE channels between the ONUs and OLTs are used to encapsulate the packets of TDM services. Symmetrical TDM PWs are created between the 2 OLTs to provide cross-PSN-network access and backhaul for enterprise or BTS private line services.

 Copyright © Huawei Technologies Co., Ltd.

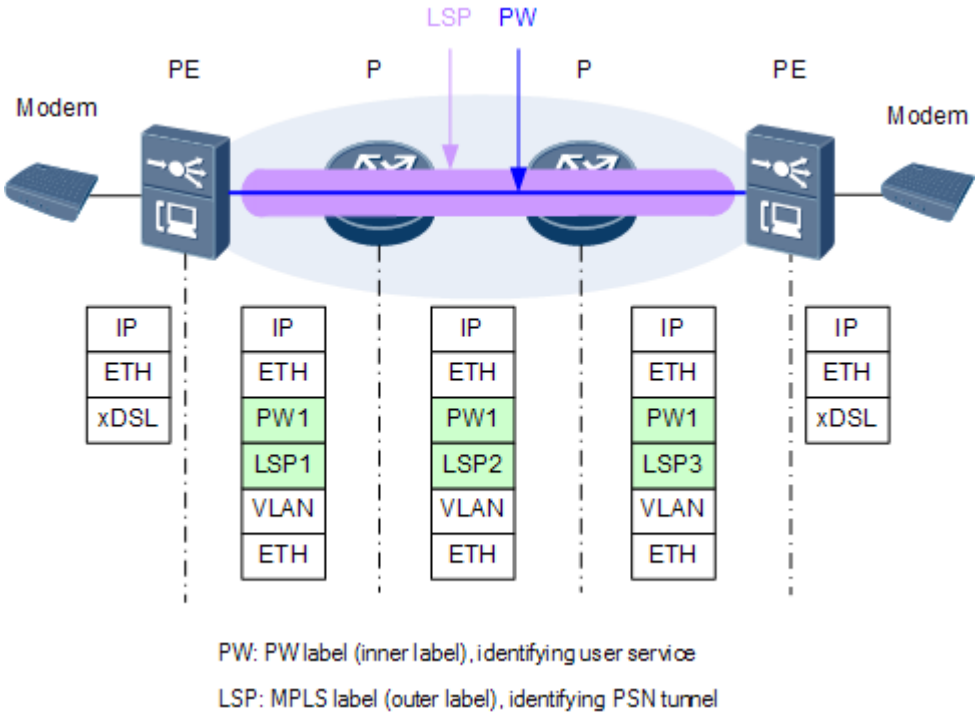
Figure 9-13 Example network of symmetrical GPON TDM PWE3 access



QoS Processing

Figure 9-14 shows the QoS processing flow of TDM PWE3 service, considering the example of SAToP (MPLS over MPLS encapsulation).

Figure 9-14 QoS processing flow of upstream SAToP service (MPLS over MPLS encapsulation)



9.1.3.3 Principle of ETH PWE3

Processing Flow of ETH PWE3 Protocol Packets

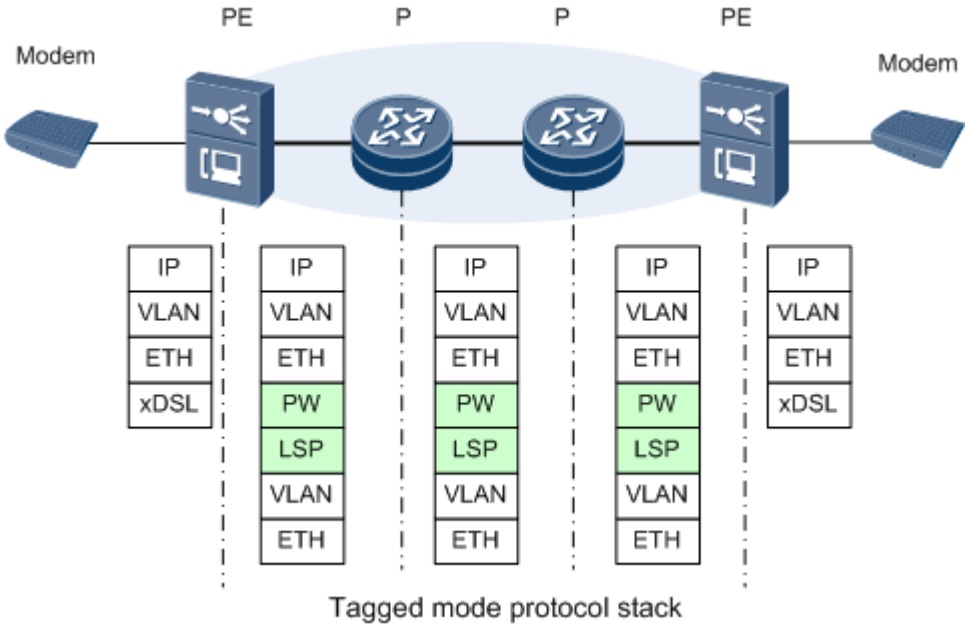
Two encapsulation formats are defined in RFC4448 for the PWE3 emulation of Ethernet service in a PSN network.

 Copyright © Huawei Technologies Co., Ltd.

- Tagged mode. In this mode, the packet going upstream carries the PW VLAN tag in the payload, and is stripped of the PW VLAN tag when going downstream.
- Raw mode. In this mode, the packet going upstream does not carry the PW VLAN tag; the PW payload, however, can carry the service VLAN tag.

Figure 9-15 shows the processing flow of ETH PWE3 protocol packets on the PE and P devices. The MA5800 can function as a PE or a P device. The PE establishes an MPLS-based PW tunnel, encapsulates the user data packets with two labels at the Tx end and transmits the packets. The P device forwards the packets. The PE at the Rx end decapsulates the received MPLS packets, restores the original user data packets, and transmits the packets to the user.

Figure 9-15 Processing flow of ETH PWE3 protocol packets

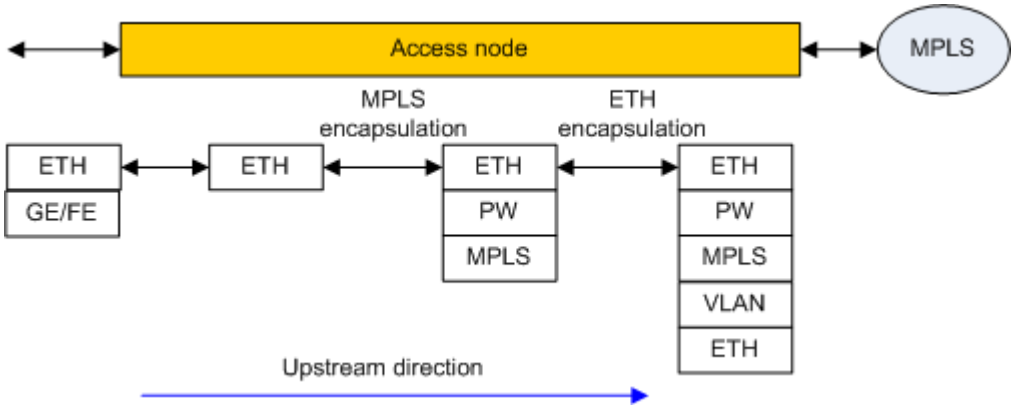


Processing Flow of ETH PWE3 Service

Figure 9-16 shows the processing flow of ETH PWE3 service (MPLS over MPLS encapsulation).

 Copyright © Huawei Technologies Co., Ltd.

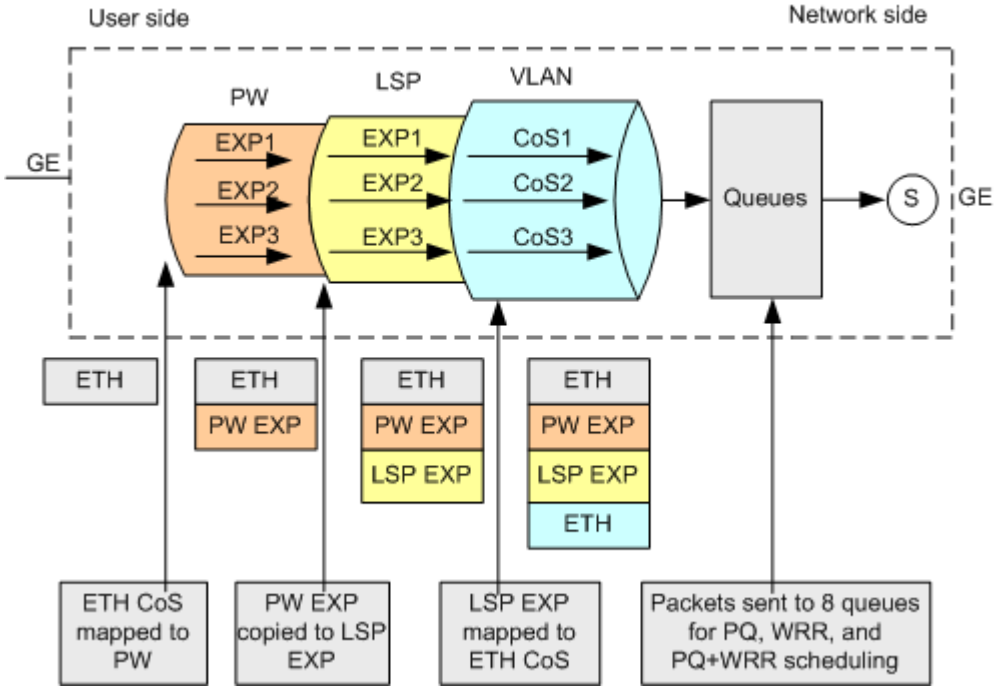
Figure 9-16 Processing flow of ETH PWE3 service



QoS Processing Flow of ETH PWE3 Service

Figure 9-17 shows the QoS processing flow of ETH PWE3 service (MPLS over MPLS encapsulation).

Figure 9-17 QoS processing flow of upstream ETH PWE3 service (MPLS over MPLS encapsulation)



9.1.3.4 Traffic Label Principle

Context

As services are developing, service traffic over a PW becomes heavier and heavier. For example, in the case of the wholesale service, certain carriers encapsulate all service traffic on hundredsor thousands of ports into a PW, and as a result traffic over the PW reaches the gigabit level. As such, a PW carrying heavy traffic is

 Copyright © Huawei Technologies Co., Ltd.

called fat PW. The fat PW burdens the equipment that it traverses, and results in congestion, packet loss, and unguaranteed QoS.

To solve problems incurred by the fat PW, IETF proposed a traffic label solution: Traffic over a fat PW takes different paths from the PW ingress PE to the PW egressPE within the network through load balancing (ECMP). To achieve PWE3 load balancing, PW data at the PW ingress PE is segmented into bundles of data streams and an MPLS label (traffic label) is allocated to every data stream. In this way, every traffic label identifies a different data stream and the traffic label is stored in the innermost area of the ingress PE label stack. Later, data is forwarded in the load balancing mode according to the traffic label, taking different data streams along different paths within the network.

In terms of PW load balancing, data over a PW arrives at the destination along different paths and this may incur disordered packets. Given this, this technology is applicable to only those services that are not sensitive to disordered packets, such as the Ethernet service. For the ATM and TDM emulation services, this technology is not applied because they have strict requirements on timing and order of packets.

PW load balancing is implemented based on equal cost LSP that is based on the equal cost route.

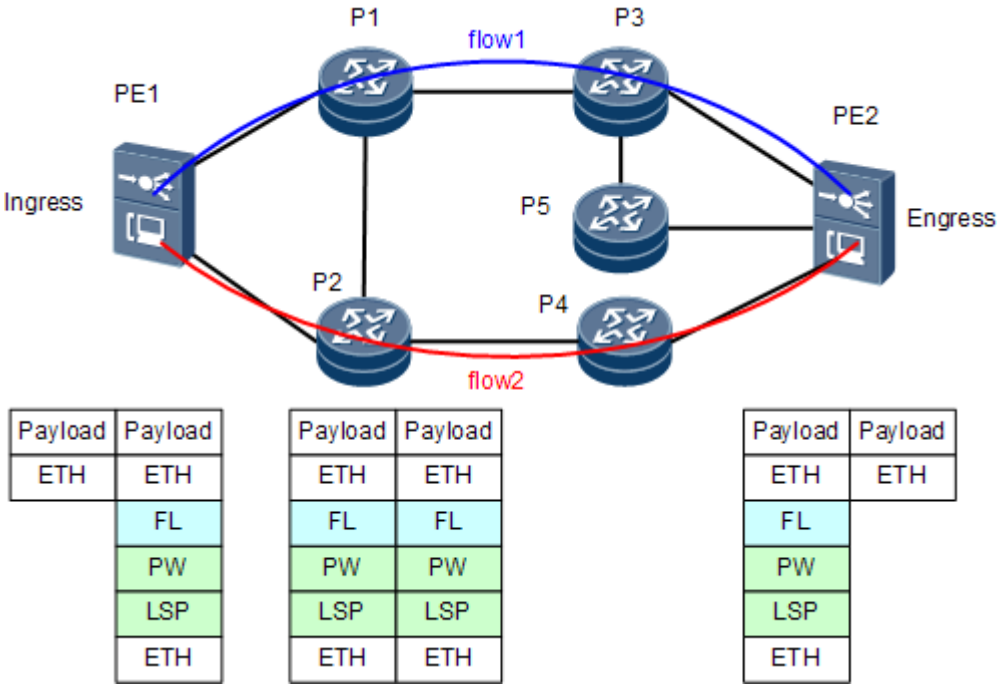
Application Description

Figure 9-18 shows an application over an existing network where a large amount of the legacy equipment does not support the traffic label. In this application, the PE supports generating the traffic label and performing load balancing but the P has no traffic label capability (for example, the legacy equipment over the network).

- PE1 generates traffic label (FL) and at the same time performs load balancing (flow1 and flow2). PE2 removes the FL.
- P1, P2, and P3 do not support traffic label for load balancing and they only forward data like a common P.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-18 Traffic label application (P equipment does not support traffic label for load balancing)



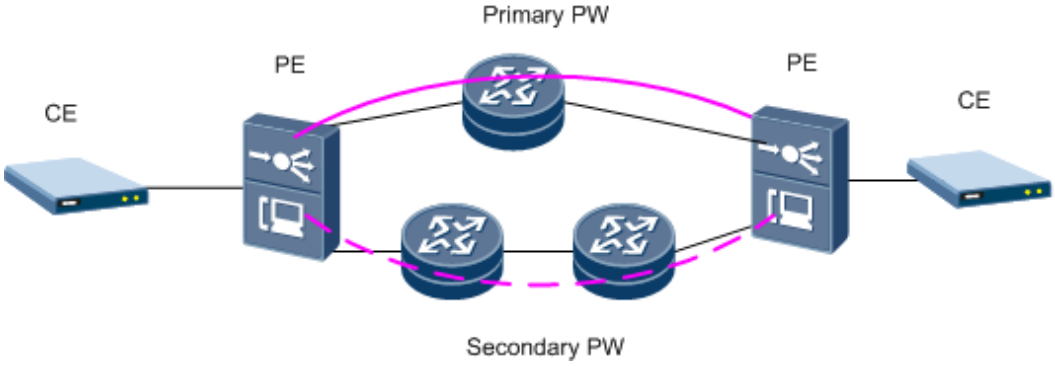
9.1.3.5 PW 1:1 Redundancy

PW redundancy is used for the PW protection switching, which has a similar function as the MPLS OAM. Different from the MPLS OAM (protecting the outer tunnel), PW redundancy protects the inner tunnel.

Through PW redundancy, data is switched to the standby PW if the active PW is faulty (such as the LDP session is down, the tunnel is deleted, the protocol communication is faulty, the route status is changed, or there is no VCCV response). In this case, the original standby PW becomes the active PW.

Figure 9-19 shows the MA5800 supports the following PW 1:1 redundancy applications.

Figure 9-19 PW 1:1 redundancy applications



 Copyright © Huawei Technologies Co., Ltd.

9.1.3.6 PW Redundancy

PW Redundancy Signaling Mechanism

Introduction of the PW protection mechanism will break the original model of 1-to-1 mapping between AC and PW in PWE3. To keep the original forwarding action, you must ensure that only one PW in the redundancy PW group is in the active state and other PWs are in the standby state.

The LDP PW signaling (RFC4447) requires to use PW status TLV to transfer the PW forwarding status. PW status TLV can be carried by the label mapping message or notification message. PW status TLV is a 32-bit status code and each bit identifies a PW forwarding state. Based on this status code, PW redundancy introduces a new PW status code (0x00000020 - PW forwarding standby) to indicate that the PW is in the standby state currently.

Primary/Secondary and Active/Inactive

There are two couples of important concepts in PW redundancy and the detailed descriptions are as follows:

- Primary/Secondary refers to the PW forwarding priority and is the PW configuration parameter.

  The primary PW is preferentially used to forward traffic and the secondary PW is used to protect primary PW. The primary PW is used to forward traffic when the state of the primary and secondary PWs is the same. Currently, only one secondary PW can be configured for each primary PW.
- Active/Inactive refers to the PW forwarding status. It indicates the PW running status and is not the configuration parameter.

  Only the PW in the active state can be used to forward traffic. The local active or inactive state of a PW is determined by the local and remote signaling status and priority (configured primary/secondary) of the PW. Only the PW in the optimal state and with the highest priority can be selected as the active PW to forward traffic, and all other PWs are in the inactive state. PWs in the inactive state are not used to forward traffic but can be enabled to receive traffic (can be used only for VLL PW).

PW Redundancy Working Mode

The PW redundancy working mode is specified on PE that is configured with active and standby PWs.

Master/Slave mode:

In this mode, the local end determines the active or standby state of the PW and uses the signaling protocol to notify the remote end; the remote PE can sense the active or standby state. The active/standby relationship on the PW side and the active/standby relationship on the AC side do not affect each other. Therefore, faults can be isolated between the PW side and the AC side. The MA5800 currently can function as a master in this mode.

Independent mode:

 Copyright © Huawei Technologies Co., Ltd.

In this mode, the active or standby state of the local PW is determined by the negotiation result of the remote AC side; the remote end notifies the local end of the active or standby state. The protection switching due to faults on the AC side will cause the protection switching on the PW side. Therefore, faults cannot be isolated.

9.1.3.7 PW OAM (VCCV)

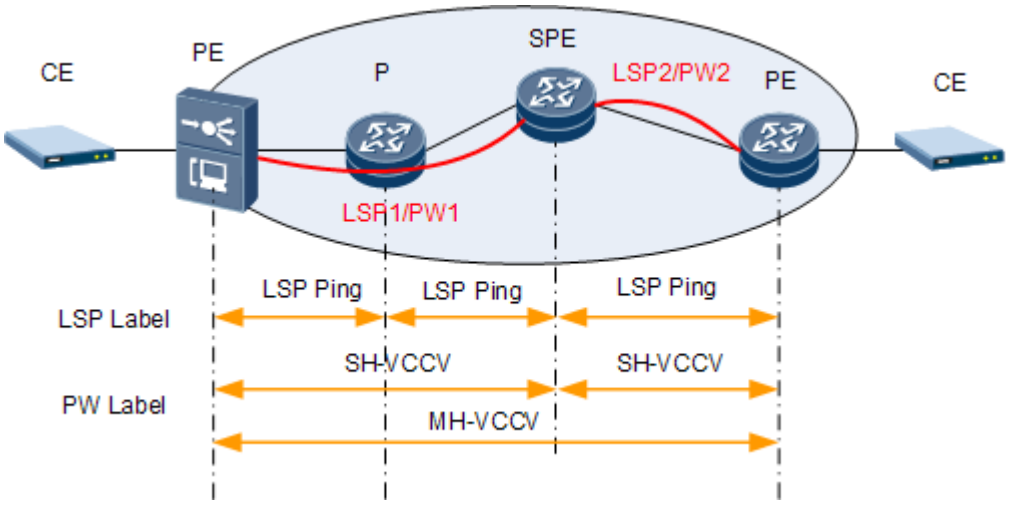
VCCV Ping Application

Virtual Circuit connectivity verification (VCCV) is an end-to-end mechanism to detect and diagnose PW faults. To put it simply, VCCV is a control channel for transmitting connectivity verification messages between PW ingress and PW egress.

VCCV ping is a tool for manually detecting the connectivity status of virtual circuits. It is implemented through extending LSP ping. VCCV defines a series of messages exchanged among PEs to verify PW connectivity. To ensure that the VCCV packet traverses the same path as the data packet in the PW, the VCCV packet must be encapsulated in the same encapsulation mode as the PW and must traverse the same tunnel as the PW packet.

VCCV ping can be used on the U-PE to detect PW connectivity, including detecting the connectivity of the static PW, dynamic PW, single-hop PW, and multi-hop PW. The MA5800 supports single-hop-VCCV (SH-VCCV) ping and does not support multi-hop-VCCV (MH-VCCV) ping temporarily. **Figure 9-20** shows the VCCV ping application.

Figure 9-20 VCCV ping application

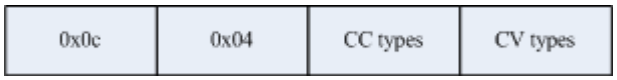


Principle

VCCV ping is implemented through the VCCV packet and the LSP ping packet therein carries the target FEC stack. **Figure 9-21** describes the parameters of the VCCV packet.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-21 VCCV packet parameters



- CC Types indicates the control channel type. **Figure 9-22** describes the CC for the VCCV function, which is defined in RFC5085.

  Figure 9-22 CC in the VCCV packet

  ```
  MPLS Control Channel (CC) Types:

  Bit (Value)     Description
  ============    ============================================
  Bit 0 (0x01) - Type 1: PWE3 Control Word with 0001b as
                 first nibble (PW-ACH, see [RFC4385])
  Bit 1 (0x02) - Type 2: MPLS Router Alert Label
  Bit 2 (0x04) - Type 3: MPLS PW Label with TTL == 1
  Bit 3 (0x08) - Reserved
  Bit 4 (0x10) - Reserved
  Bit 5 (0x20) - Reserved
  Bit 6 (0x40) - Reserved
  Bit 7 (0x80) - Reserved
  ```

  - Type1: control word channel. Whether VCCV is performed depends on the control word (0001 or not). SH-VCCV and MH-VCCV are supported. If the PE supports control word, CC type1 is used preferentially.
  - Type2: MPLS router alert channel. Whether VCCV is performed depends on a specific label value (label = 2). SH-VCCV is supported and MH-VCCV is not supported.
  - Type3: maximum-hop channel. VCCV is performed if the TTL value of the inner label of MPLS is 1. SH-VCCV and MH-VCCV are supported.
- CV Types indicates the connectivity verification type. **Figure 9-23** describes the CV for the VCCV function, which is defined in RFC5085. The MA5800 supports only CV of the LSP ping type.

  Figure 9-23 CV in the VCCV packet

  ```
  MPLS Connectivity Verification (CV) Types:

  Bit (Value)     Description
  ============    ============================================
  Bit 0 (0x01) - ICMP Ping
  Bit 1 (0x02) - LSP Ping
  Bit 2 (0x04) - Reserved
  Bit 3 (0x08) - Reserved
  Bit 4 (0x10) - Reserved
  Bit 5 (0x20) - Reserved
  ```

T-PE peers at both sides negotiate CC and CV capabilities during PW set-up, and then send the same CC and CV types used by VCCV ping as the negotiation result. If the PE supports control word, CC type1 is used preferentially. SH-VCCV and MH-VCCV are implemented through setting different inner PW label TTL values.

Figure 9-24 shows the CC Type1 VCCV flow.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-24 CC Type1 MH-VCCV/SH-VCCV flow

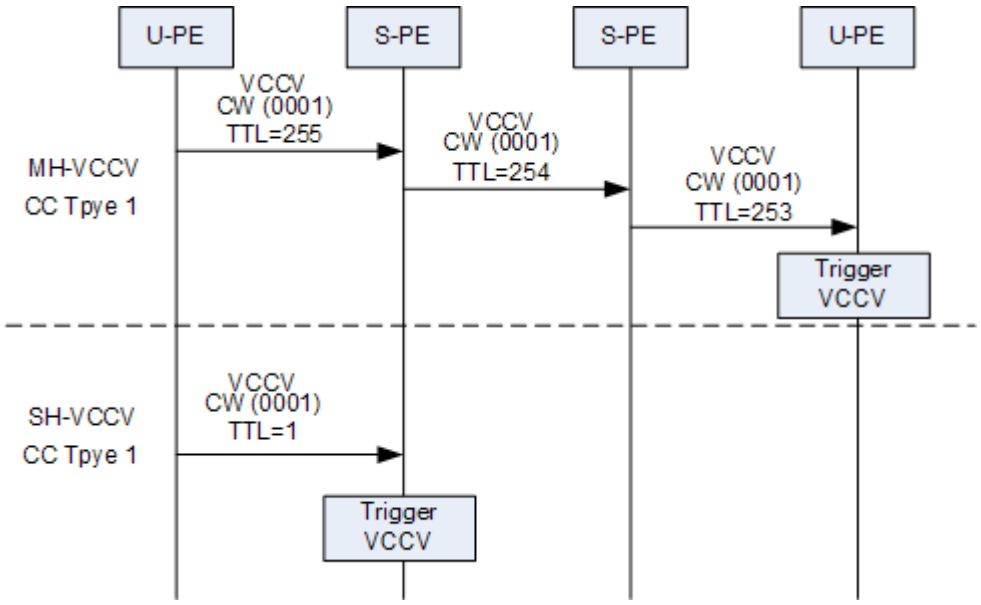
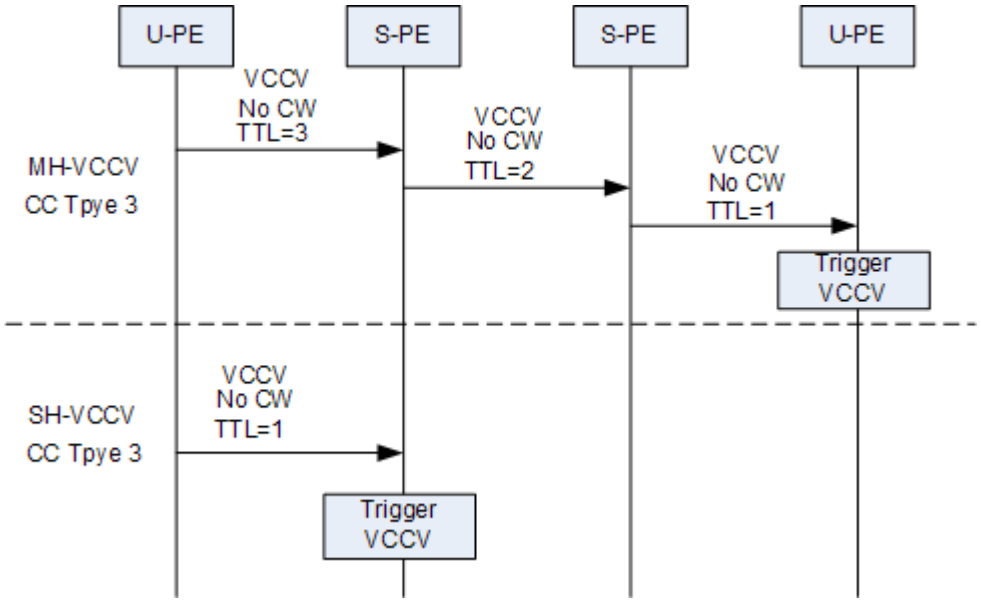


Figure 9-25 shows the CC Type3 VCCV flow.

Figure 9-25 CC Type3 MH-VCCV/SH-VCCV flow



The U-PE and S-PE differ in processing the VCCV packet:

- The S-PE pays attention to only the PW TTL value. If the PW TTL value is 1, VCCV is performed.
- The U-PE pays attention to not only CW (VCCV is performed if the first half-byte is 0001) but the PW TTL value (VCCV is performed if the PW TTL value is 1).
- If the U-PE initiates VCCV to the S-PE, a proper TTL value needs to be set, irrespective of whether CW is used.

 Copyright © Huawei Technologies Co., Ltd.

9.1.4 PWE3 Network Applications

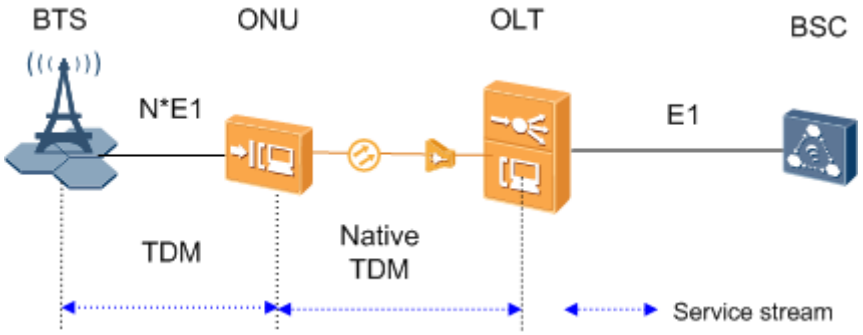
Network Applications of TDM PWE3

For details on the corresponding clock solution of the TDM service, see **14.3 Physical Layer Clock/Time Synchronization**.

- Network application: converting the native TDM service into the E1 service for upstream transmission

  As shown in **Figure 9-26**, the mobile 2G base station is connected to the ONU through TDM E1. The ONU encapsulates the TDM frame into the GPON GEM frame directly to transmit TDM service over the GPON network (native TDM). The OLT terminates the TDM PWE3 signals and transmits the signals to the upstream network through the E1 port.

  Figure 9-26 Network application: converting the TDM PWE3 service into the E1 service for upstream transmission

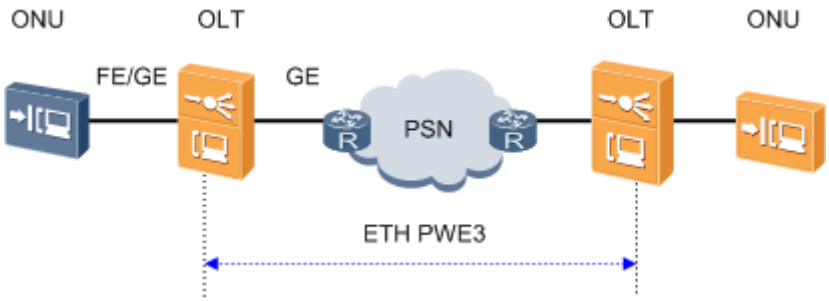


Network Applications of ETH PWE3

- Network application: implementing PWE3 private line upstream transmission in FE/GE access

  As shown in **Figure 9-27**, the enterprise router is connected to the OLT through FE/GE. The OLT interconnects with the peer ETH PWE3 device of the enterprise through the ETH PWE3 private line to implement the ETH private line service.

  Figure 9-27 Network application: implementing PWE3 private line upstream transmission in FE/GE access



Copyright © Huawei Technologies Co., Ltd.

9.1.5 Configuring the PWE3 Private Line Service

Pseudo wire emulation edge-to-edge (PWE3) uses LDP or RSVP-TE as the signaling protocol and carries various Layer 2 services of the customer edge (CE) over the MPLS LSP or TE tunnel, transparently transmitting the Layer 2 data of the CE.

PWE3 Service Model

According to the PWE3 service model, PWE3 is indicated by the outer packet switch network (PSN) tunnel label and the inner label (PW demultiplexer).

Network Application

Figure 9-28 shows the network application of the MPLS PWE3.

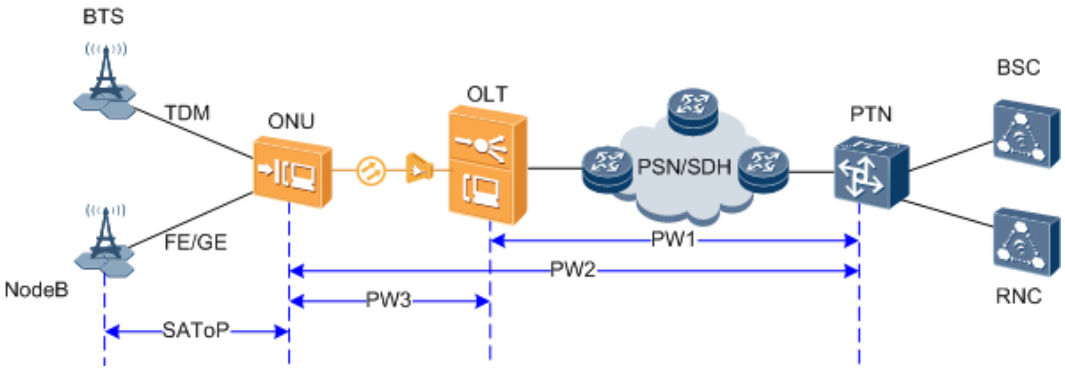
As shown in the figure, the mainstream applications of the MPLS PWE3 supported by the MA5800 are as follows:

- TDM PWE3: A mobile 2G base station is connected to the ONU through the TDM E1 port. The ONU implements the TDM PWE3, transmitting traffic streams to the peer TDM PWE3 device through the PSN. The MA5800 functions as a Layer 2 transparent transmission device, PE device, or P device.
- ETH PWE3: A 3G base station is connected to the ONU through the FE/GE port. The ONU performs the ETH PWE3 encapsulation for interconnecting with the peer ETH PWE3 device. The MA5800 functions as a Layer 2 transparent transmission device or P device.

NOTE

- The MA5800 can function as a Layer 2 transparent transmission device or PE/P device, determined by the service requirement.
- As shown in the following figure, in PW1 and PW3, the MA5800 functions as a PE device that initiates or terminates the PW; in PW2, the MA5800 functions as a Layer 2 transparent transmission device or P device.

Figure 9-28 MPLS PWE3 network application when the MA5800 functions as a Layer 2 transparent transmission device, PE device, or P device



 Copyright © Huawei Technologies Co., Ltd.

Procedure

According to the PWE3 service model, PWE3 configurations include the outer tunnel configuration, inner PW configuration, and tunnel protection. Therefore, the configuration procedure is as follows.

9.1.5.1 Configuring the PWE3 Outer Tunnel

To provide services across the PSN network, the MA5800 supports PW over the MPLS tunnel to transparently transmit services in the PSN network.

Prerequisites

1. The loopback interface IP address must be configured.
2. The LSR ID must be configured.
3. The global MPLS and MPLS TE functions must be enabled.
4. The host route of each port must be successfully advertised on each device in the network (For example, the OSPF protocol was configured).

Context

Only the MPLS LSP is supported as PWE3 outer tunnel currently.

Procedure

- Configure the MPLS TE tunnel.
  a. In the global config mode, run the **interface tunnel** command to create a tunnel interface and enter the tunnel interface mode.
  b. Run the **tunnel-protocol mpls te** command to configure the tunnel protocol to MPLS TE, that is, configure the tunnel interface to work in the TE tunnel mode.
  c. Run the **destination ip-address** command to configure the destination IP address of the tunnel. Generally, the LSR ID of the ingress is used.
  d. Run the **mpls te tunnel-id** command to configure the tunnel ID.
  e. Run the **mpls te signal-protocol { cr-static | rsvp-te }** command to configure the signaling protocol for the MPLS TE tunnel.

     According to whether the MPLS TE tunnel uses the dynamic signaling protocol, the tunnel is categorized as create te tunnel based on static cr-lsp and MPLS RSVP-TE tunnel.

     - Create te tunnel based on static cr-lsp : The forwarding information and resource information are configured manually, and the signaling protocol and path calculation are not involved. Because the MPLS-related control packets are not exchanged, fewer resources are used. The static tunnel, however, cannot be dynamically adjusted according to network changes. Therefore, the actual application is limited.
     - MPLS RSVP-TE tunnel: MPLS TE creates the LSP tunnel along a specified path through RSVP-TE and reserves resources. Thus, carriers can accurately control the path that traffic traverses to avoid the node where congestion occurs. This solves the problem that certain

 Copyright © Huawei Technologies Co., Ltd.

paths are overloaded and other paths are idle, utilizing the current bandwidth resources sufficiently.

f. (Optional) Run the **mpls te bandwidth** command to configure the bandwidth for the tunnel. After the configuration is completed, only the VLAN interface meeting this bandwidth requirement is selected as the node traversed by an MPLS TE tunnel when the MPLS TE tunnel is created.

If the MPLS TE tunnel is only used to change the data transmission path, you may not configure the bandwidth of the tunnel.

g. (Optional) Run the **mpls te path explicit-path** command to configure the explicit path used by the MPLS TE tunnel.

To limit only the bandwidth of the MPLS TE tunnel but not the transmission path, you may not configure the explicit path of the tunnel.

h. Run the **display interface tunnel** command to query the configuration of the tunnel.

- For the configuration of LDP LSP, see **6.7.2 Configuring the LDP LSP**.

----End

9.1.5.2 Configuring the Tunnel Policy

Configure the tunnel selection sequence for load balancing or the tunnel binding policy in the tunnel. After the configuration is successful, packets in the tunnel are processed according to tunnel policy.

Prerequisites

The PWE3 outer tunnel must be created.

Context

The tunnel selection sequence and the tunnel binding policy are mutually exclusive. This means that you can configure only one of them.

- The IP tunnel supports the configuration of only the tunnel selection sequence.
- The MPLS TE tunnel supports the configuration of only the tunnel binding policy.

Procedure

Step 1 Run the **tunnel-policy** command to create a tunnel policy name and enter the tunnel policy mode.

Step 2 For IP tunnel, run the **tunnel select-seq** command to configure the selection sequence of tunnels for load balancing.

To configure different tunnel types for load balancing according to priorities, run this command. The more the tunnel type close to keyword **select-seq**, the higher priority for load balancing.

The MA5800 does not support load balancing between different tunnels. In other words, tunnels for load balancing must be of the same type. The tunnels are selected according to the tunnel configuration.

 Copyright © Huawei Technologies Co., Ltd.

Step 3 For MPLS TE tunnel, run the **tunnel binding** command to configure the tunnel binding policy.

To bind to a specified tunnel ID and configure the system to switch another tunnel according to the configured sequence when a tunnel is not available, run this command. After the tunnel binding policy is configured, run the **mpls te reserved-for-binding** command in the tunnel mode to allow the MPLS TE tunnel to be bound to the VPN instance.

destination *ip-addr* indicates the destination IP address of the tunnel, which must be the same as the destination IP address configured in the MPLS TE tunnel.

Step 4 In the global config mode, run the **display tunnel-policy** command to query the information about the tunnel policy.

----End

Example

To configure a tunnel policy named te_policy and bind to tunnels with the destination IP address 10.5.5.5 and IDs 10 and 20, do as follows:

```
huawei(config)#tunnel-policy te_policy
Info: New tunnel-policy is configured.
huawei(config-tunnel-policy-te_policy)#tunnel binding destination 10.5.5.5 te tunnel
 10 tunnel 20
huawei(config)#display tunnel-policy
{ <cr>|string<S><Length 1-19> }:

  Command:
        display tunnel-policy
Total   tunnel policy num:             1
Sel-Seq tunnel policy num:              0
Binding tunnel policy num:              1
Invalid tunnel policy num:              0

Tunnel Policy Name  Destination     Tunnel Intf                 Down switch
------------------------------------------------------------------------------
te_policy           10.5.5.5        tunnel10                    Disable
                                    tunnel20
```

9.1.5.3 Configuring the PWE3 Inner PW

Configure the attribute of PW and use the PW parameters for PW binding.

Prerequisites

- MPLS L2VPN must be enabled.
- The outer tunnel must be configured.

Context

PW parameters include the following parameters: control word, jitter buffer (only for TDM PWs), maximum transmission unit (MTU), loopback IP address of the peer device, PW type, RTP control header, virtual circuit connectivity verification (VCCV), used tunnel policy, flow label classification, and TDM load time (only for TDM PWs). Different services have different configurations when the services are bound to a PW.

 Copyright © Huawei Technologies Co., Ltd.

Procedure

Step 1 Run the **pw-para** command to create PW parameter.

PW parameters and the PW have a one-to-one mapping. One PW parameter can be used by only one PW.

Step 2 Run the **peer-address** command to configure the IP address of the peer device.

peer-address indicates the peer IP address in the PW for creating communication. In the actual transmission, data packets are automatically transmitted to the peer device according to this IP address.

Step 3 Run the **pw-type** command to configure the PW type.

The MA5800 supports TDM and ETH PWs.

ETH PWs are categorized as raw and tagged modes.

- Raw mode: The PW VLAN tag is not carried in the upstream direction, but the PW payload can carry the SVLAN.
- Tagged mode: The payload of an upstream packet carries the PW VLAN tag, and the PW VLAN tag is removed in the downstream direction.

For the same PW, the PW types at both ends must be the same. In this way, the PW can be available.

NOTICE

Among PW parameters, the IP address and PW type of the peer device cannot be changed after they are configured. To change these two parameters, run the **undo pw-para** command to delete them first, and then configure them again. Make sure that the two parameters are correctly configured the first time, so as to prevent repeated operations.

Step 4 (Optional) Run the **control-word** command to enable the control word mode.

When VCCV ping works in the control word mode, you need to enable the control word. It is recommended that you enable the control word mode.

Step 5 (Optional) Run the **jitter-buffer** command to configure the jitter buffer.

The jitter buffer can effectively prevent jitter and delay. By default, the jitter buffer size is 2000 μs.

NOTE

- Only a TDM PW supports setting of the jitter buffer size.
- The jitter buffer size must be an integer multiple of 125.

Step 6 (Optional) Run the **mtu** command to configure the MTU.

Due to the limit in the system, the configurable MTU ranges for different PW types are different:

- MTU values set on the two devices at the ends of an ETH PW must be the same. If MTU values are different, an ETH PW can never be available.
- By default, the MTU is 1500 bytes. Do not modify this value unless there is a special requirement.

 Copyright © Huawei Technologies Co., Ltd.

Step 7 Run the **rtp-header** command to configure the RTP control header.

NOTE

This command is applicable to only TDM PWs.

The length of the RTP header is 12 bytes, including the version number, padding flag, and timestamp fields. The timestamp field, whose length is 32 bits, is used for clock synchronization. For format of the RTP header, see RFC3550.

After RTP is enabled, PW packets of the TDM type carry the RTP control header. Otherwise, the RTP control header is not carried.

The RTP configuration must be the same as that on the peer PW device. By default, the MA5800 disables the RTP control header.

Step 8 (Optional) Run the **vccv** command to enable VCCV, so as to notify the peer device of the VCCV types supported by the local device. After a successful negotiation between both devices, a virtual circuit connectivity verification is performed by using LSP ping according to the priority of the VCCV type.

VCCV is an end-to-end PW fault detection and diagnosis mechanism. Simply, VCCV is a control channel for the PW to send verification messages between the ingress and egress.

Enable the LSP ping function for alter, CW, and TTL channels or any of the three channels according to the VCCV types supported by the system. By default, VCCV is disabled.

Step 9 (Optional) Run the **tdm-load-time** command to configure the TDM load time.

NOTE

Only a TDM PW supports the setting of the load time.

Because each TDM frame is 125 μs, the load time must be an integer multiple of 125. If the entered number is not an integer multiple of 125, the system rounds it down to the nearest integer multiple of 125 μs. The jitter buffer must be greater than the load time.

The default jitter buffer is 1000 μs. Do not modify this value unless there is a special requirement.

Step 10 (Optional) Run the **tnl-policy** command to configure the tunnel policy used by the PW.

NOTE

The tunnel policy and the PW flow label classification are mutually exclusive. Configure either of them.

After the tunnel policy used by the PW is configured, the PW can perform load balancing or path selection according to the tunnel policy.

Step 11 In the privilege mode or global config mode, run the **display pw-para** command to query the configuration of the PW.

----End

Example

To configure PW 10 with the following attributes, do as follows:

 Copyright © Huawei Technologies Co., Ltd.

- IP address of the peer PW device: 10.10.10.20
- PW type: TDM CESOPSN
- Name of the tunnel policy used by the PW: **tdm-policy**
- Enable the RTP control header and the control word mode
- Enable the connectivity verification function of the alter, CW and TTL channels
- Other parameters: default settings

```
huawei(config)#pw-para pwindex 10
huawei(config-pw-para-index-10)#service-type vll
huawei(config-pw-para-index-10)#pwid 10
huawei(config-pw-para-index-10)#peer-address 10.10.10.20
huawei(config-pw-para-index-10)#pw-type tdm cesopsn
huawei(config-pw-para-index-10)#tnl-policy tdm-policy
huawei(config-pw-para-index-10)#rtp-header
huawei(config-pw-para-index-10)#control-word
huawei(config-pw-para-index-10)#vccv cc cw alert ttl cv lsp-ping
huawei(config-pw-para-index-10)#quit
huawei(config)#display pw-para pwindex 10
  PW index          : 10
  Service type      : VLL
  PW ID             : 10
  PeerIP            : 10.10.10.20
  Tnl Policy Name   : --
  PW Type           : tdm satop e1
  Remove Double Tag : --
  Tag Mode          : --
  Cos Group ID      : --
  Pbits-to-pbits ID : --
  CtrlWord          : enable
  VCCV capability   : cw alert ttl/lsp-ping
  BFD compatibility : disable
  MTU               : 1500
  Statistic switch  : disable
  Dynamic in label  : --
  MaxAtmCells       : --
  MaxEncapDelay     : --
  RTP               : enable
  PayloadType       : 0
  SnStartNum        : 0
  RcvSsrcId         : 0
  TxSsrcId          : 0
  JitterBuffer      : 2000
  LoadTime(us)      : 1000
  TimeSlotNum       : 32
  PayLoadSize(bytes): 256
  FlowLabel transmit           : --
  FlowLabel classification-rule : --
  FlowLabel receive            : --
  Priority mapping profile name : default-profile-2
```

To configure PW 20 with the following attributes, do as follows:

- IP address of the peer PW device: 10.20.30.40
- PW type: ETH Tagged
- Name of the tunnel policy used by the PW: **eth-policy**
- Other parameters: default settings

```
huawei(config)#pw-para pwindex 20
huawei(config-pw-para-index-20)#service-type vll
huawei(config-pw-para-index-20)#pwid 20
huawei(config-pw-para-index-20)#peer-address 10.20.30.40
huawei(config-pw-para-index-20)#pw-type ethernet tagged
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-pw-para-index-20)#tnl-policy eth-policy
huawei(config-pw-para-index-20)#quit
huawei(config)#display pw-para pwindex 20
  PW index          : 20
  Service type      : VLL
  PW ID             : 20
  PeerIP            : 10.20.30.40
  Tnl Policy Name   : --
  PW Type           : ethernet tagged
  Remove Double Tag : --
  Tag Mode          : --
  Cos Group ID      : 0
  Pbits-to-pbits ID : --
  CtrlWord          : disable
  VCCV capability   : disable
  BFD compatibility : disable
  MTU               : 1500
  Statistic switch  : disable
  Dynamic in label  : --
  MaxAtmCells       : --
  MaxEncapDelay     : --
  RTP               : --
  PayloadType       : --
  SnStartNum        : --
  RcvSsrcId         : --
  TxSsrcId          : --
  JitterBuffer      : --
  LoadTime(us)      : --
  TimeSlotNum       : --
  PayLoadSize(bytes): --
  FlowLabel transmit           : disable
  FlowLabel classification-rule : --
  FlowLabel receive            : disable
  Priority mapping profile name : default-profile-0
```

9.1.5.4 Binding the Service to the PW

Bind various PWE3 services to a PW. After the binding, user packets are encapsulated and forwarded according to the modes defined in the PW parameters.

Prerequisites

- The PW must be configured.
- For TDM PWE3, the TDM connection must be created.
- For ETH PWE3, the ETH-based service port must be created.

Context

Only the dynamic PW is supported currently.

Dynamic PW is negotiated using the LDP protocol.

Procedure

- Bind the TDM service to a PW.

  Run the **pw-ac-binding tdm** command to use a PW to create the TDM PW service.

  Pay attention to the following points during the configuration:

  – To specify a PW as a static PW, you need to configure the in label and out label of the PW. The out label value must be an unallocated and idle

 Copyright © Huawei Technologies Co., Ltd.

value at the peer end and the in label value must be an unallocated value at the local end.

- Bind the ETH service to a PW.

  Run the **pw-ac-binding vlan** command to use a PW to create the ETH PW service.

----End

Example

To create a dynamic binding between TDM connection 0 and PW 20, do as follows:

```
huawei(config)#pw-ac-binding tdm 0 pw pwindex 20
```

To bind the ETH service to a PW with the following settings, do as follows. Settings: VLAN ID 100, PW ID 30.

```
huawei(config)#pw-ac-binding vlan 100 pw pwindex 30
```

9.1.5.5 Configuring PW Protection

Create a standby PW for a PW. When the active PW is faulty, the system quickly switches to the standby PW to ensure the service reliability.

Prerequisites

- The active PW must be created.
- The basic parameters of secondary PW are configured. For the configuration method, see **9.1.5.3 Configuring the PWE3 Inner PW**.

Context

PW protection: When a PW is faulty (such as an LDP session is down, a tunnel is deleted, the protocol communication is faulty, the route status changes, or VCCV has no response), the system can quickly switch to the standby PW. Then, the standby PW functions as the active PW.

The MA5800 supports PW 1:1 redundancy.

Procedure

Step 1 Run the **pw-protect** command to configure the standby PW.

Pay attention to the following points during the configuration:

- The standby PW ID cannot exist.
- The PW parameters of the active and standby PWs must be the same.
- Both active PW and the standby PW are not static PW.

Step 2 (Optional) Enable the PW protection group to support dual-sending and dual-receiving for the multicast service.

When the PW protection group supports dual-sending, both active and standby PWs can forward IGMP packets so that the multicast forwarding entry can also be created on the device corresponding to the standby PW. After the active/standby

 Copyright © Huawei Technologies Co., Ltd.

PW switchover, the multicast service can be smoothly switched. This configuration is recommended when the multicast service is carried by the active and standby PWs.

When the PW protection group supports dual-receiving, both active and standby PWs can receive packets to avoid packet loss caused by signaling delay when switchback is performed after the faulty active PW recovers. This configuration is recommended when the multicast service is carried by the active and standby PWs.

1. Run the **igmp_send_dual-pw** command to set whether IGMP packets can be sent by both active and standby PWs.
2. Run the **pw-redundancy_stream-dual-receiving** command to set the PW protection group to work in the dual-receiving mode.

Step 3 Run the **pw-revertive-mode** command to configure the switchback policy for the PW protection group.

Switchback: When both active and standby PWs are available, if the original service traffic is carried on the standby PW, the service can be switched back to the active PW according to actual requirements. Set the switchback policy according to actual network conditions (such as whether the network topology often changes and whether the traffic should be carried on the active PW).

The switchback policy of a PW protection group can be immediate automatic switchback, automatic switchback after a period of time, and no automatic switchback.

Step 4 Run the **display pw-ps** command to query the configuration of the PW protection group.

----End

Example

To configure a PW protection group, set the parameters as follows: active PW ID to 10, standby PW ID to 20, and switchback policy to allowing automatic switchback for the PW protection group in 30 seconds.

```
huawei(config)#pw-protect primary-pw pwindex 10 secondary-pw pwindex 20
huawei(config)#pw-revertive-mode pwindex 10 revertive wtr 30
huawei(config)#display pw-ps pwindex 10
  ------------------------------------------------------------------------------
  Primary-PW-Index  Primary-PW-state    Secondary-PW-Index   Secondary-PW-state
  ------------------------------------------------------------------------------
             10             down                 20               down
  ------------------------------------------------------------------------------
  stream-dual-receiving: disable
  pw-redundancy mode: master
  pw-redundancy switchover config: primary
  revertive-mode: revertive, in 30 seconds
```

9.1.5.6 Configuring MPLS Tunnel Protection

Create a protection tunnel for the MPLS TE tunnel. When the working tunnel is faulty, the system quickly switches to the protection tunnel to ensure the service reliability.

 Copyright © Huawei Technologies Co., Ltd.

Prerequisites

- The forward LSP must be created.
- The backward LSP must be created.
- MPLS OAM must be enabled.

Context

MPLS tunnel protection is a part of the MPLS OAM connectivity detection mechanism.

The basic process of the MPLS OAM connectivity check and protection switching is as follows:

1. The source transmits the CV/FFD packets to the destination through the detected LSP.
2. The destination checks the correctness of the type and frequency carried in the received detection packets and measures the number of correct and incorrect packets that are received within the detection period to monitor the connectivity of the LSP in real time.
3. After detecting a defect, the destination transmits the BDI packets that carry the defect information to the source through the backward path.
4. The source learns about the status of the defect, and triggers the corresponding protection switching when the protect group is correctly configured.

Procedure

Step 1 Configure working MPLS TE tunnel.

1. In the global config mode, run the **interface tunnel** command to create a tunnel interface and enter the tunnel interface mode.
2. Run the **tunnel-protocol mpls te** command to configure the tunnel protocol to MPLS TE.
3. Run the **destination ip-address** command to configure the destination IP address of the tunnel. Generally, the egress LSR ID is used.
4. Run the **mpls te tunnel-id** command to configure the tunnel ID.
5. Run the **mpls te signal-protocol rsvp-te** command to configure the signaling protocol of the tunnel to RSVP-TE.
6. (Optional) Run the **mpls te bandwidth** command to configure the bandwidth for the tunnel. After the configuration is completed, only the VLAN interface that meets this bandwidth value can be selected as the node traversed by the MPLS TE tunnel path when the MPLS TE tunnel is created.

   If the MPLS TE tunnel is only used to change the data transmission path, you may not configure the tunnel bandwidth.
7. (Optional) Run the **mpls te path explicit-path** command to configure the explicit path used by the MPLS TE tunnel.

   If only the bandwidth used by the MPLS TE tunnel is limited but the transmission path is not limited, you may not configure the explicit path used by the MPLS TE tunnel.

 Copyright © Huawei Technologies Co., Ltd.

Step 2 Configure protection MPLS TE tunnel.

The working mode of MPLS OAM protection switching is 1:1 protection. Normally, each working tunnel has a protection tunnel.

The configuration of the protection tunnel is the same as that of the working tunnel.

Step 3 Configure a tunnel protect group.

Configure the working tunnel and the protection tunnel as a tunnel protect group. When the source end finds the active LSP is defective through the MPLS OAM detection mechanism, and the protection switching is required, the system can switch the data to the protection tunnel for continuous transmission.

1. In the global config mode, run the **interface tunnel** command to enter the working tunnel interface mode.
2. Run the **mpls te protection tunnel** command to create a tunnel protect group and set the switchback mode of the protect group.

   The switchback policy of a PW protection group can be immediate automatic switchback, automatic switchback after a period of time, and no automatic switchback.

Step 4 (Optional) Run the **mpls te protect-switch** command forcibly switch over the tunnel protect group.

To manually switch data streams between working and protection tunnels, run this command.

There are for forcible switching modes:

- **clear**: clears all external switching commands that are already executed in the system.
- **lock**: lock switching, which locks data streams on the working tunnel.
- **force**: forcible switching, which forcibly switches data streams to the protect tunnel.
- **manual work-lsp**: manually switches data streams on the working tunnel to the protection tunnel.
- **manual protect-lsp**: manually switches data streams on the protection tunnel to the working tunnel.

Keywords **clear**, **lock**, **force**, and **manual** corresponds to switching priorities in descending order. If a command with a higher priority is executed, a command with a lower priority cannot be executed.

Step 5 In the global config mode, run the **display mpls te protection tunnel** command to query the configuration of the tunnel protect group.

----End

Example

To configure RSVP-TE tunnel IDs to 10 and 30, destination IP address of the tunnels to 10.3.3.3, tunnel 30 as the protection tunnel of tunnel 10, switchback mode to revertive, and WTR time to 900s, do as follows:

```
huawei(config)#interface tunnel 10
huawei(config-if-tunnel10)#tunnel-protocol mpls te
huawei(config-if-tunnel10)#destination 10.3.3.3
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-if-tunnel10)#mpls te tunnel-id 10
huawei(config-if-tunnel10)#mpls te signal-protocol rsvp-te
huawei(config-if-tunnel10)#mpls te bandwidth ct0 5120  //(Optional) Configure the global bandwidth of tunnel 10 to 5120 kbit/s.
huawei(config-if-tunnel10)#mpls te commit
huawei(config-if-tunnel10)#quit
huawei(config)#interface tunnel 30
huawei(config-if-tunnel30)#tunnel-protocol mpls te
huawei(config-if-tunnel30)#destination 10.3.3.3
huawei(config-if-tunnel30)#mpls te tunnel-id 30
huawei(config-if-tunnel30)#mpls te signal-protocol rsvp-te
huawei(config-if-tunnel30)#mpls te bandwidth ct0 5120  //(Optional) Configure the global bandwidth of tunnel 30 to 5120 kbit/s.
huawei(config-if-tunnel30)#mpls te commit
huawei(config-if-tunnel30)#quit
huawei(config)#interface tunnel 10
huawei(config-if-tunnel10)#mpls te protection tunnel 30 mode revertive wtr 30
huawei(config-if-tunnel10)#mpls te commit
huawei(config-if-tunnel10)#quit
```

9.1.5.7 Configuring CR-LSP Backup

Backup CR-LSPs are established on networks requiring high reliability to provide end-to-end protection, ensuring network reliability. If a primary CR-LSP fails, traffic rapidly switches to a backup CR-LSP, ensuring uninterrupted traffic transmission.

Prerequisites

Before configuring CR-LSP backup, complete the following tasks:

- Setting up a primary RSVP-TE tunnel
- Enabling MPLS, MPLS TE, and RSVP-TE globally and on interfaces of each node along a backup CR-LSP

Context

Hot standby and ordinary backup modes are supported. If both primary and backup CR-LSPs fail, best-effort paths are established. The following table lists CR-LSP backup modes.

Backup Mode	Description	Advantage	Shortcoming
Hot standby	A hot-standby CR-LSP is set up over a separate path immediately after a primary CR-LSP is set up	A rapid traffic switchover can be performed	Additional bandwidth needs to be reserved for a hot-standby CR-LSP
Ordinary backup	The system attempts to set up an ordinary backup CR-LSP if a primary CR-LSP fails	No additional bandwidth is needed	Ordinary backup performs a traffic switchover slower than hot standby

 Copyright © Huawei Technologies Co., Ltd.

Backup Mode	Description	Advantage	Shortcoming
Best-effort path	The system establishes a best-effort path over an available path if both the primary and backup CR-LSPs fail	Establishing a best-effort path is easy and a few constraints are needed	Some quality of service (QoS) requirements cannot be met

The following backup modes are supported for CR-LSP.

- Hot standby
- Ordinary backup
- Hot standby + best-effort path

Procedure

Step 1 Configure the backup mode of CR-LSP.

1. Run the **interface tunnel** command to enter MPLS TE tunnel mode.
2. Run the **mpls te backup** command to configure the backup mode of current tunnel.
   - The keyword **hot-standby** indicates the hot standby mode.
   - **ordinary** indicates the ordinary mode.
   - **ordinary best-effort** indicates the best-effort path mode.

Step 2 (Optional) Configure the explicit path for a backup CR-LSP.

After hot standby or ordinary backup is configured, the system automatically selects a path for a backup CR-LSP. To manually specify a path for a backup CR-LSP, you can set explicit path. An explicit path consists of a series of nodes, which constitute a vector path according to the configured sequence. The IP address in an explicit path is the IP address of the interface on the node. Generally, the loopback interface IP address on the egress is used as the destination IP address of the explicit path.

NOTE

Use a separate explicit path for the backup CR-LSP to prevent the backup CR-LSP from completely overlapping its primary CR-LSP. Protection will fail if the backup CR-LSP completely overlaps its primary CR-LSP.

1. Run the **explicit-path** command in the global config mode to create an explicit path.
2. Run the **next hop**, **modify hop**, and **delete hop** command to add a next hop node, modify a node, and delete a node respectively for the explicit path.
3. In the tunnel mode, run the **mpls te path explicit-path** *path-name* **secondary** command to specify the explicit path for the backup CR-LSP.

Step 3 (Optional) Configure the affinity property for a backup CR-LSP.

 Copyright © Huawei Technologies Co., Ltd.

Affinity property masks determine the link properties that should be checked by a device. To ensure that a link can be used by a tunnel, for the bits that are 1 in a mask, it is required that at least one bit in the administrative group and the corresponding bit in the affinity property be 1. In addition, if the bits in the affinity property are 0, the corresponding bits in the administrative group cannot be 1.

The default affinity property is 0x0.

Run the **mpls te affinity property secondary** command to set the affinity property for the backup CR-LSP.

Step 4 (Optional) Configure the hop limit for a backup CR-LSP.

After a CR-LSP is configured with a hop limit, the hop limit acts as one of routing conditions such as the link bandwidth and affinity property when the CR-LSP is created. After the hop limit is set, the number of hops of a CR-LSP cannot exceed this limit.

The default hop limit is 32.

Run the **mpls te hop-limit secondary** command to set hop limit for the backup CR-LSP.

----End

Follow-up Procedure

After the configuration of CR-LSP backup is finished, you can query information about the tunnel interface and backup status.

- Run the **display mpls te tunnel-interface** command to check information about a tunnel interface.
- Run the **display mpls te hot-standby state** command to check information about the hot-standby status.
- Run the **display mpls te tunnel** command to check CR-LSP information.

9.2 Native TDM

In Native TDM, TDM frames are directly encapsulated to GPON GEM frames. This mode features simple encapsulation, small network cost, and guaranteed link quality. TDMoGEM mode and TDMoETH mode are encapsulated to NativeTDM in access network at present, and the MA5800 supports TDMoETH mode.

9.2.1 Introduction to Native TDM

Definition

By using the standard 8 kHz (125μm) frames, the GPON GTC layer is synchronous in nature. Therefore, GPON can support the TDM service. This is called Native TDM.

In Native TDM, TDM frames are encapsulated to ETH frames and then to GPON GEM frames in TDMoETH mode. This mode features simple encapsulation, small network cost, and guaranteed link quality.

 Copyright © Huawei Technologies Co., Ltd.

Purpose

Currently, the circuit switched network is evolving to the packet switched network. During the deployment of the packet switched network, the method to provide traditional circuit switching service over the packet switched network must be considered. In a GPON network deployment, the traditional TDM service can be delivered over the PSTN network through the Native TDM mode.

9.2.2 Native TDM Reference Standards and Protocols

The following lists the reference documents of Native TDM:

- ITU-T G.984.1 General characteristics for Gigabit-capable Passive Optical Networks (GPON)
- ITU-T G.984.2 Gigabit-capable Passive Optical Networks (GPON): Physical Media Dependent (PMD) layer specification
- ITU-T G.984.3 Gigabit-capable Passive Optical Networks (GPON): Transmission convergence layer
- ITU-T G.984.4 Gigabit-capable Passive Optical Networks (GPON): ONT management and control interface specification

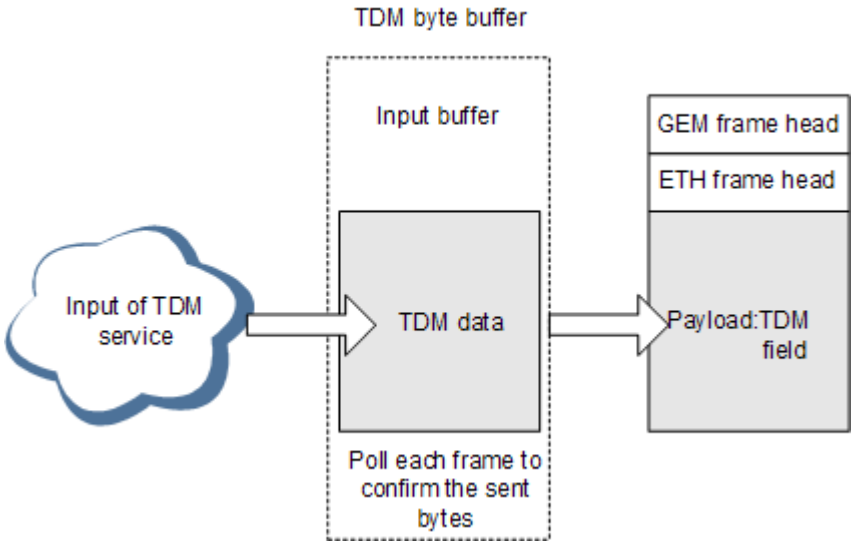
9.2.3 Native TDM Principle

Basic Principle

The MA5800 supports E1 upstream transmission mode. The following section uses E1 upstream transmission as an example.

In the TDMoETH mode, the TDM frame is directly encapsulated to the GPON GEM frame so that the TDM service can be transmitted in a GPON network. **Figure 9-29** shows the basic principle of Native TDM.

Figure 9-29 Basic principle of TDMoETH mode Native TDM



 Copyright © Huawei Technologies Co., Ltd.

The TDMoETH mode does not distinguish whether it is a structured service such as the voice and PBX access services, or an un-structured service such as private line service. All services are handled as un-structured services. GPON access transparently delivers only the E1 service without performing the refined service processing. In other words, GPON access provides the long-haul E1 transmission.

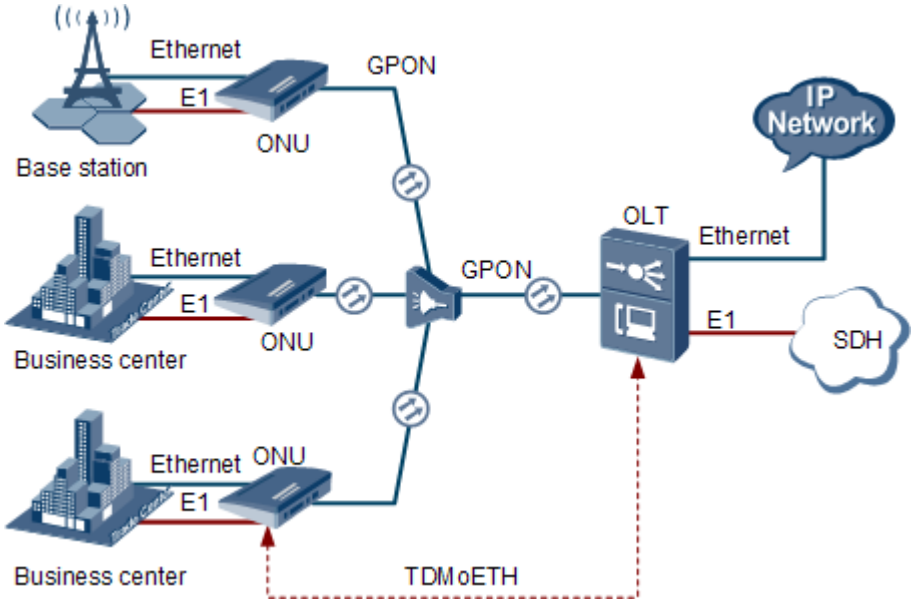
In general, the private line service has an independent clock, which is asynchronous with the GPON line clock. The jitter and frequency difference exist between the two clocks. Therefore, the E1 service traffic can be transmitted only after the E1 rate adjustment occurs at the ONT/ONU.

The bit asynchronous mapping mode of SDH is adopted while the GPON line clock or system clock is used as the reference clock. Through the bit adjustment, the E1 service traffic is mapped to the payload section of the tributary unit to form a standard VC12 virtual channel, which is delivered in the GEM frame. At the receiver end, de-byte adjustment is adopted to recover the original clock and the E1 service traffic.

Application

Figure 9-30 shows the TDMoETH network application. The E1 ports of the ONU/ONT access the TDM traffic from the base stations and enterprises. The ONU/ONT sends both the Ethernet traffic and the TDM traffic to the OLT. The OLT differentiates the Ethernet traffic and the TDM traffic, and sends the traffic to the IP network and the SDH network, respectively.

Figure 9-30 TDMoETH mode Native TDM network application



9.3 VPLS

The Virtual Private LAN Service (VPLS), also called the Transparent LAN Service (TLS) or virtual private switched network service, is a Layer 2 VPN (L2VPN)

 Copyright © Huawei Technologies Co., Ltd.

technology that is based on Multi-Protocol Label Switching (MPLS) and Ethernet technologies.

9.3.1 What Is VPLS

Definition

The Virtual Private LAN Service (VPLS), also called the Transparent LAN Service (TLS) or virtual private switched network service, is a Layer 2 VPN (L2VPN) technology that is based on Multi-Protocol Label Switching (MPLS) and Ethernet technologies.

Purpose

The primary goal of VPLS is to interconnect multiple Ethernet LANs through the Packet Switched Network (PSN). In this manner, these LANs can function as one LAN. VPLS can implement the multipoint-to-multipoint VPN networking; therefore, by using the VPLS technology, service providers (SPs) can provide the Ethernet-based multipoint services through MPLS backbone networks. In addition, by utilizing the VPLS solution in which MPLS virtual circuits (VCs) function as the Ethernet bridge links, SPs can transparently transmit LAN services on the MPLS network.

9.3.2 VPLS Reference Standards and Protocols

The following table lists the references of this document.

Document No.	Description
RFC 4762	Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling
draft-ietf-l2vpn-oam-req-frmk-01	VPLS OAM Requirements and Framework
RFC 4447	Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)
draft-ietf-l2vpn-signaling-08	Provisioning, Autodiscovery, and Signaling in L2VPNs
draft-ietf-pwe3-redundancy-04	Pseudowire Redundancy

9.3.3 VPLS Principles

9.3.3.1 Introduction to VPLS

Basic VPLS Transport Structure

Figure 9-31 shows an example of a VPLS network. The entire VPLS network is similar to a switch. PWs are established over MPLS tunnels between VPN sites to

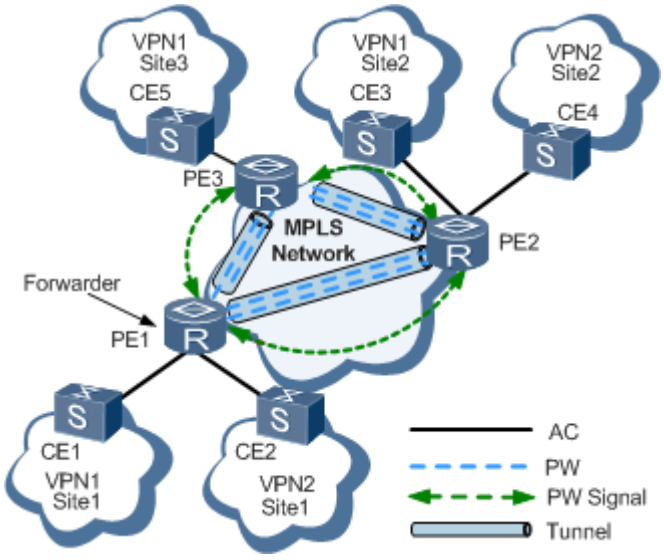
 Copyright © Huawei Technologies Co., Ltd.

transparently transmit Layer 2 packets between sites. When forwarding packets, PEs learn the source MAC addresses of these packets and create MAC entries, mapping MAC addresses to attachment circuits (ACs) and PWs.

The following table describes the various concepts related to VPLS networks.

- AC A link between a CE and a PE. An AC must be established using Ethernet interfaces. On a VPLS network, AC interfaces can be Ethernet interfaces, Ethernet sub-interfaces, VLANIF interfaces, Eth-Trunk interfaces, Eth-Trunk sub-interfaces, VE interfaces, QinQ interfaces, and VE (ATM 1483B) interfaces.
- PW A bidirectional virtual connection between two virtual switch instances (VSIs) residing on two PEs. A PW consists of a pair of unidirectional MPLS VCs transmitting in opposite directions.
- VSI A type of instance used to map ACs to PWs. A VSI independently provides VPLS services and forwards Layer 2 packets based on MAC addresses and VLAN tags. A VSI has the Ethernet bridge function and can terminate PWs.
- PW signaling A type of signaling used to create and maintain PWs. PW signaling is the foundation for VPLS implementation. Currently, the PW signaling is LDP or BGP. MA5800 supports only LDP PW signaling.
- Tunnels are used for carrying PWs. A tunnel can carry multiple PWs. A tunnel is a direct channel that transparently transmits data between the local and remote PE devices. It can be an MPLS or a GRE tunnel. MA5800 supports only MPLS tunnels.
- Forwarder Similar to a VPLS forwarding table. After a PE receives packets from an AC, the forwarder of the PE selects a PW to forward these packets. It is similar to a VPLS forwarding table.

Figure 9-31 Basic VPLS transmission process



The forwarding of a packet from CE1 to CE3 on VPN1 is used as an example:

1. CE1 sends a Layer 2 packet to PE1 over an AC.
2. After PE1 receives the packet, the forwarder of PE1 selects a PW for forwarding the packet.

 Copyright © Huawei Technologies Co., Ltd.

3. PE1 then adds two MPLS labels to the packet based on the PW forwarding entry and sends the packet to PE2. The private network label identifies the PW, and the public network label identifies the tunnel between PE1 and PE2.
4. After PE2 receives the packet from the public tunnel, PE2 removes the private network label of the packet.
5. The forwarder of PE2 selects an AC and forwards the packet to CE3 over the AC.

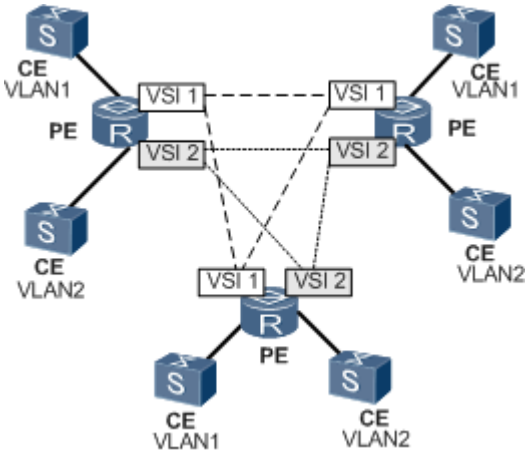
VPLS Implementation Process

Transmission of packets between CEs relies on VSIs configured on PEs, and PWs established between the VSIs. **Figure 9-31** shows transmission of Ethernet frames over full-mesh PWs between PEs.

The Ethernet often uses the Spanning Tree Protocol (STP) to prevent loops. VPLS networks, however, use full-mesh PWs and split horizon to avoid loops as follows:

The PEs in a VSI must be fully meshed. That is, a PE must create a tree path to every other PE in the VSI. Each PE must support split horizon to avoid loops. Split horizon requires that packets received from a PW in a VSI should not be forwarded to other PWs in the VSI. Any two PEs in a VSI must communicate over a direct PW, which is why full-mesh PWs are required between PEs in a VSI.

Figure 9-32 VPLS forwarding model



A VPLS network consists of a control plane and a forwarding plane.

- The control plane of a VPLS PE provides the following functions:
  - Member discovery: a process in which a PE in a VSI discovers the other PEs in the same VSI. This process can be implemented manually or automatically using protocols. BGP VPLS and BGP AD VPLS both support automatic member discovery.
  - Signaling mechanism: PWs between PEs in the same VSI are established, maintained, or torn down using signaling protocols such as LDP and BGP.
- The forwarding plane of a VPLS PE provides the following functions:
  - Encapsulation: After receiving Ethernet frames from a CE, a PE encapsulates the frames into packets and sends the packets to a PSN.

 Copyright © Huawei Technologies Co., Ltd.

- Forwarding: A PE determines how to forward a packet based on the inbound interface and destination MAC address of the packet. For details, see **Table 9-3**.
- Decapsulation: After receiving packets from a PSN, a PE decapsulates these packets into Ethernet frames and sends the frames to a CE.

VPLS Implementation Modes

VPLS can be implemented in LDP, BGP, or BGP AD mode.

- VPLS implemented in LDP mode is also called Martini VPLS.
- VPLS implemented in BGP mode is also called Kompella VPLS.
- VPLS BGP AD uses extended BGP Update packets to implement automatic member discovery. It also uses LDP FEC 129 signaling packets for local and remote VSIs to automatically negotiate and establish VPLS PWs.

NOTE

MA5800 only supports LDP mode.

The differences between the three tunnel setup modes are as follows:

- In LDP tunnel setup mode, the requirements for PEs are low, but no auto-discovery mechanism for VPN members can be provided, which has to be configured manually. In BGP tunnel setup mode, the requirements for PEs are high. That is, PEs must run BGP. In addition, the auto-discovery mechanism for VPN members can be provided. Therefore, if the LDP mode is used, you need to configure the PW to the new PE on each PE when adding a PE. In BGP mode, if the number of PEs does not exceed the label block size, you only need to configure new PEs.
- In LDP tunnel setup mode, an LDP session must be created between every two PEs. The number of sessions is in direct ratio to the square of the number of PEs. In BGP tunnel setup mode, route reflector (RR) can be used to reduce the number of BGP connections.
- In LDP tunnel setup mode, each PE is assigned with a label only if necessary. In BGP tunnel setup mode, each PE is assigned with a label block, which leads to the waste of labels.
- In LDP tunnel setup mode, the VSIs configured in all domains must use the same VSI ID range. In BGP tunnel setup mode, the VPN target is used to identify VPNs.
- The BGP AD VPLS mode integrates the advantages of the Kompella and Martini VPLS modes. With support for automatic VPLS member discovery and PW deployment, BGP AD VPLS simplifies VPLS configurations on newly added nodes as compared with Martini VPLS. BGP AD VPLS also saves label resources and is compatible with PWE3, which are not supported by Kompella VPLS.

Table 9-2 shows the comparison between the two VPLS tunnel setup modes.

 Copyright © Huawei Technologies Co., Ltd.

Table 9-2 Comparison between two VPLS tunnel setup modes

Type	LDP	BGP
Requirements for PEs	Common	High
Auto-discovery supported	No	Yes
Implementation complexity	Low	High
Expansibility	Poor	Good
Label utilization ratio	High	Low
Configuration workload	High	Low
Cross-domain restrictions	High	Low

After the preceding comparison, the following conclusions can be drawn:

- The LDP tunnel setup mode is preferable when the number of VPLS sites is relatively small, the VPLS network seldom or never traverses multiple domains, and PEs do not run BGP.
- The BGP tunnel setup mode is applicable at the core layer of a large-scale network when PEs run BGP and cross-domain is required.

If the scale of a VPLS network is large (a great number of nodes in a wide geographical range), you can use HVPLS to combine the two modes. That is, the core layer uses the BGP tunnel setup mode and the access layer uses the LDP tunnel setup mode. VPLS assumes that each PE is capable of setting up tunnels; PW labels functions as the identifiers for services; tunnels are responsible for transmitting VPLS data from a PE to another PE.

VPLS Encapsulation Modes

- Packet encapsulation on ACs

  Packet encapsulation on ACs depends on the user access mode, which can be VLAN or Ethernet access.

  NOTE

  Currently, the MA5800 supports only packet encapsulation type of VLAN.

  Table 9-3 Packet encapsulation on ACs

Packet Encapsulation Type	Description
VLAN	The header of each Ethernet frame sent between CEs and PEs carries a VLAN tag, known as the SVLAN. This is a service delimiter identifying users on an ISP network.

 Copyright © Huawei Technologies Co., Ltd.

Packet Encapsulation Type	Description
Ethernet	The header of each Ethernet frame sent between CEs and PEs does not carry an SVLAN. If the frame header contains a VLAN tag, it is an inner VLAN tag called the CVLAN. A CE does not add the CVLAN to an Ethernet frame; instead, the tag is carried in a packet before the packet is sent to the CE. A CVLAN informs the CE to which VLAN the packet belongs, and is meaningless to PEs.

- Packet encapsulation on PWs

  The PW ID and PW encapsulation type uniquely identify a PW. The PW IDs and PW encapsulation types configured on the two end PEs of a PW must be the same. The packet encapsulation types of packets on PWs can be raw or tagged. By default, packets on PWs are encapsulated in tagged mode.

  Table 9-4 Packet encapsulation on PWs

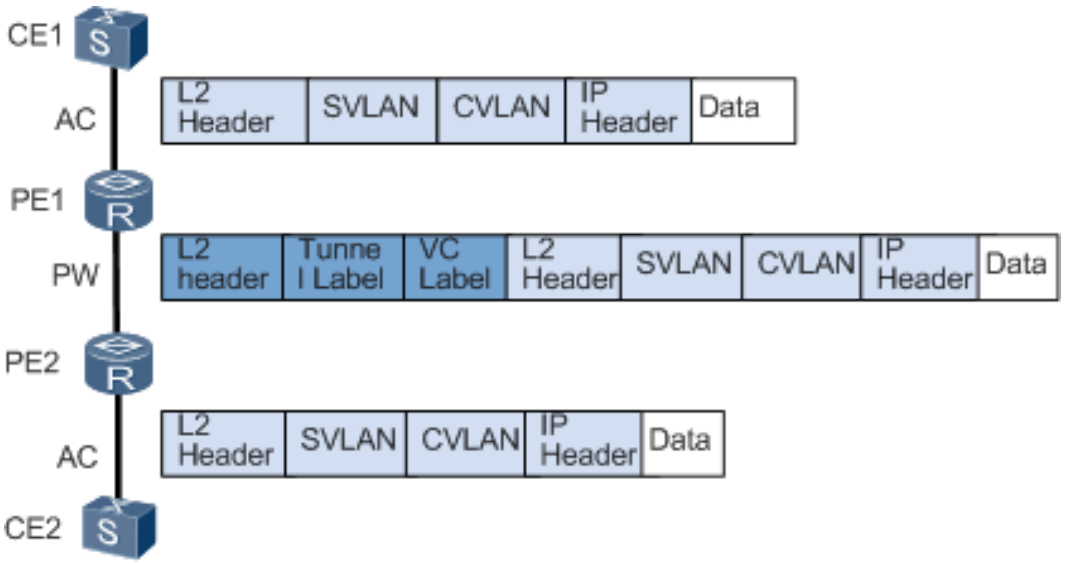
Packet Encapsulation Type	Description
Raw	Packets transmitted over a PW cannot carry SVLANs. If a PE receives a packet with the SVLAN from a CE, the PE strips the SVLAN and adds double labels (outer tunnel label and inner VC label) to the packet before forwarding it. If a PE receives a packet with no SVLAN from a CE, the PE directly adds double labels (outer tunnel label and inner VC label) to the packet before forwarding it. The PE determines whether to add the SVLAN to a packet based on actual configurations before sending it to a CE. The PE is not allowed to rewrite or remove an existing CVLAN.
Tagged	Packets transmitted over a PW must carry SVLANs. If a PE receives a packet with the SVLAN from a CE, the PE directly adds double labels (outer tunnel label and inner VC label) to the packet before forwarding it. If a PE receives a packet with no SVLAN from a CE, the PE adds a null SVLAN and double labels (outer tunnel label and inner VC label) to the packet before forwarding it. The PE determines whether to rewrite, remove, or preserve the SVLAN of a packet based on actual configurations before forwarding it to a CE.

Encapsulation modes of packets transmitted over ACs and PWs can be used together. The following uses VLAN+tagged encapsulation (with the CVLAN) as examples to describe the packet exchange process.

VLAN+tagged encapsulation (with the CVLAN)

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-33 VLAN+tagged encapsulation (with the CVLAN)



As shown in **Figure 9-33**, ACs use VLAN encapsulation and PWs use tagged encapsulation; packets transmitted from CEs to PEs carry U-Tags and SVLANs.

The packet exchange process is as follows:

1. CE1 sends a packet that has Layer 2 encapsulation and carries both a CVLAN and an SVLAN to PE1.
2. Upon receiving the packet that contains the CVLAN and SVLAN, PE1 does not process the CVLAN and regards it as user data.
3. The PW is encapsulated in tagged mode. A frame to be sent to the PW must carry the SVLAN. After receiving the packet with the SVLAN, PE1 does not process the packet.
4. PE1 queries entries in the VSI, and selects a tunnel and a PW for the packet.
5. PE1 adds double MPLS labels (outer tunnel label and inner VC label) to the packet based on the selected tunnel and PW, performs Layer 2 encapsulation, and forwards the packet to PE2.
6. PE2 receives the packet from PE1 and decapsulates the packet to remove Layer 2 encapsulation information and two MPLS labels.
7. PE2 forwards the original Layer 2 packet from CE1 to CE2. The packet carries the C-VLAN and S-VLAN tags.

The processing of sending a packet from CE2 to CE1 is similar to this process.

Derivative VPLS Functions

Traffic Statistics

Traffic statistics can be collected based on ACs or PWs, and the status of various types of traffic can be viewed in real time.

VPLS Service Isolation

 Copyright © Huawei Technologies Co., Ltd.

VPLS service isolation allows you to prohibit communication between users that use the same service and bound to the same VSI.

By default, traffic can be forwarded between AC interfaces, between UPE PWs, and between AC interfaces and UPE PWs in a VSI. After VPLS forwarding isolation is enabled, traffic cannot be forwarded between AC interfaces, between UPE PWs, or between an AC and a UPE PW in a VSI.

NOTE

On a non-hierarchical VPLS network, VPLS service isolation prohibits traffic forwarding between AC interfaces. On an HVPLS network, VPLS service isolation prohibits traffic forwarding between AC interfaces, between UPE PWs, and between AC interfaces and UPE PWs.

9.3.3.2 VPLS Layer 2 Functions

Background

A characteristic of the Ethernet is that a port sends unicast packets with unknown destination MAC addresses, broadcast packets, and multicast packets to all other ports on the Ethernet. As an Ethernet-based technology, VPLS emulates an Ethernet bridge for user networks. To forward packets on a VPLS network, PEs must establish MAC address tables and forward packets based on MAC addresses or MAC addresses and VLAN tags.

On a VPLS network, each VSI maintains a MAC address forwarding table and provides the functions of MAC address learning, flooding, forwarding, and aging to forward packets.

Related Concepts

- MAC address learning

  Table 9-5 describes MAC address learning modes.

  NOTE

  At present, the MA5800 supports only MAC address learning in qualified mode.

  Table 9-5 MAC address learning modes

MAC Address Learning Mode	Description	Characteristic
Qualified	A PE learns the MAC addresses and VLAN tags of received Ethernet frames. In this mode, each user VLAN is an independent broadcast domain and has independent MAC address space.	The broadcast domain is confined to each user VLAN. Qualified learning can result in large FIB table sizes, because the logical MAC address is now a VLAN tag + MAC address.

 Copyright © Huawei Technologies Co., Ltd.

MAC Address Learning Mode	Description	Characteristic
Unqualified	A PE learns only the MAC addresses of Ethernet frames. In this mode, all user VLANs share the same broadcast domain and MAC address space. The MAC address of each user VLAN must be unique.	If an AC interface is associated with multiple user VLANs, this AC interface must be a physical interface bound to a unique VSI.

- MAC address aging

  An aging mechanism removes MAC entries that a PE no longer needs. If a MAC entry is not updated within a specified period of time, this entry will be aged.

Implementation

PEs establish MAC address tables based on dynamic MAC address learning and associates destination MAC addresses with PWs. **Table 9-6** describes the MAC address learning process.

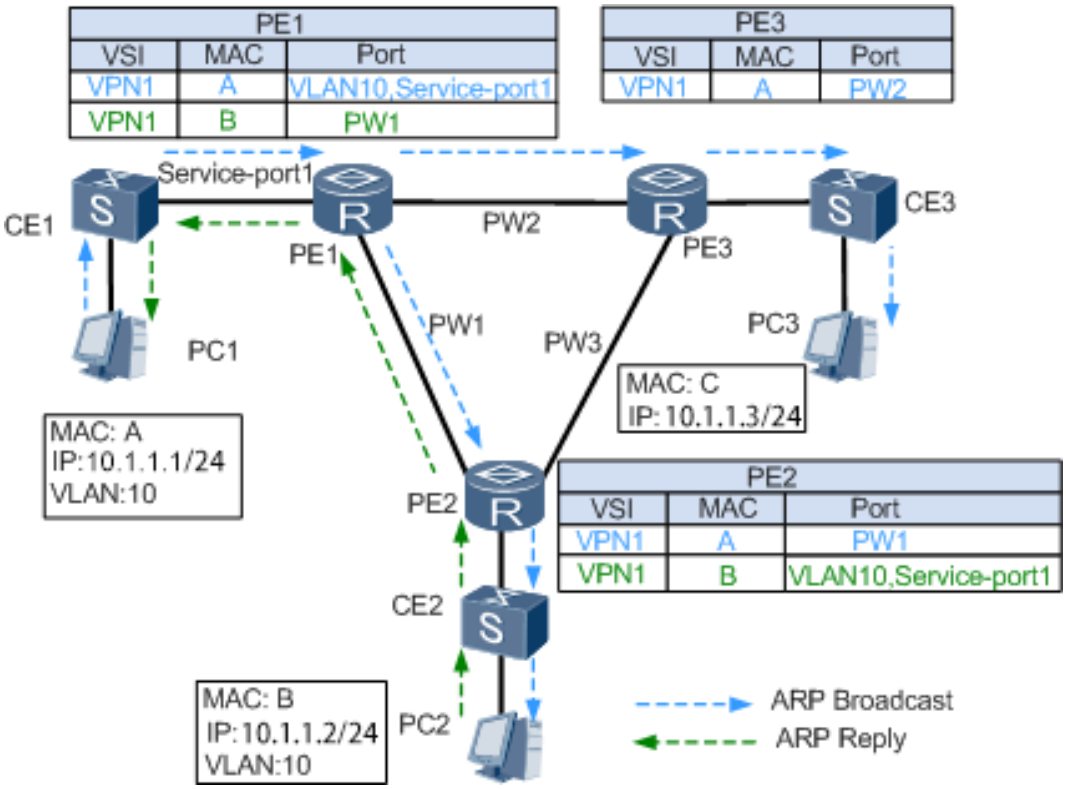
Table 9-6 MAC address learning process

MAC Address Learning Process	Description
Learning MAC addresses from user-side packets	After receiving packets from a CE, a PE maps their source MAC addresses to the service port corresponding to CE.
Learning MAC addresses from PW-side packets	A PW consists of a pair of MPLS VCs transmitting in opposite directions. A PW will go Up only after the two MPLS VCs are established. After a PE receives a packet with an unknown source MAC address from a PW, the PE maps the source MAC address to the AC interface receiving the packet.

Figure 9-34 shows the process of MAC address learning and flooding on a PE. PC1 and PC2 both belong to VLAN10. When PC1 pings IP address 10.1.1.2, PC1 does not know the MAC address corresponding to this IP address and advertises an ARP Request packet.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-34 MAC address learning process



1. After receiving the ARP Request packet sent by PC1 from service port1 that connects to CE1, PE1 adds the MAC address of PC1 to its own MAC address table, as shown in the blue section of the MAC entry.
2. PE1 advertises the ARP Request packet to its other ports (PW1 and PW2 can be viewed as ports).
3. After receiving the ARP Request packet from PW1, PE2 adds the MAC address of PC1 to its own MAC address table, as shown in the blue section of the MAC entry.
4. Based on split horizon, PE2 sends the ARP Request packet to only the port connecting to CE2 (as indicated by the blue dashed line), but not to PW1. This ensures that only PC2 receives the ARP Request packet. VPLS split horizon ensures that packets received from public network PWs are forwarded to only private networks, not to other public network PWs.
5. After PC2 receives the ARP Request packet and finds that it is the destination of this packet, PC2 sends an ARP Reply packet to PC1 (as indicated by the green dashed line).
6. After receiving the ARP Reply packet from PC2, PE2 adds the MAC address of PC2 to its own MAC address table, as indicated by the blue section of the MAC entry. The destination MAC address of the ARP Reply packet is the MAC address of PC1 (MAC A). After searching its MAC address table, PE2 sends the ARP Reply packet to PE1 over PW1.
7. After receiving the ARP Reply packet from PE2, PE1 adds the MAC address of PC2 to its own MAC address table. After searching its MAC address table, PE1 sends the ARP Reply packet to PC1 through service port1.

 Copyright © Huawei Technologies Co., Ltd.

8. After receiving the ARP Reply packet from PC2, PC1 completes MAC address learning.
9. While advertising the ARP Request packet to PW1, PE1 also advertises the ARP Request packet to PE3 over PW2. After receiving the ARP Request packet, PE3 adds the MAC address of PC1 to its MAC address table. Based on split horizon, PE3 sends the ARP Request packet to only PC3. Because PC3 is not the destination of the ARP Request packet, PC3 does not send any ARP Reply packet.

Derivative Functions

Traffic Restriction

On a VPLS network, you can limit the rates of broadcast, multicast, and unknown unicast packets to:

- Enhance traffic management and appropriately allocate user bandwidth.
- Prevent traffic attacks and enhance network security.

Limit on the Number of Learned MAC Addresses

After the number of MAC entries or MAC address learning time reaches the set threshold, a device forwards or drops newly received packets and decides whether to report an alarm to the network management system (NMS).

This function applies to networks with relatively fixed users but insufficient security, such as residential access networks and enterprise intranets without security management.

9.3.3.3 LDP VPLS

Background

LDP VPLS (Martini VPLS) uses a static discovery mechanism to discover VPLS members using LDP signaling. VPLS information is carried in extended TLV fields of LDP signaling packets.

Related Concepts

LDP VPLS involves the following concepts:

- FEC: A set of packets with similar or identical characteristics and forwarded in the same way by LSRs. Characteristics determining the FEC of a packet include the destination address, service type, and QoS attribute. Currently, the MA5800 only supports VLAN as FEC.
- TLV: A highly efficient and expansible coding mode for protocol packets. To support new features, you only need to add new types of TLVs to carry information required by the features.
- DU: A label distribution mode in which an LSR distributes labels to FECs without having to receive Label Request messages from its upstream LSR.
- Liberal: A label retention mode in which an LSR retains the label mapping received from a neighboring LSR, regardless of whether the neighboring LSR is its next hop. In liberal label retention mode, an LSR can use the labels sent

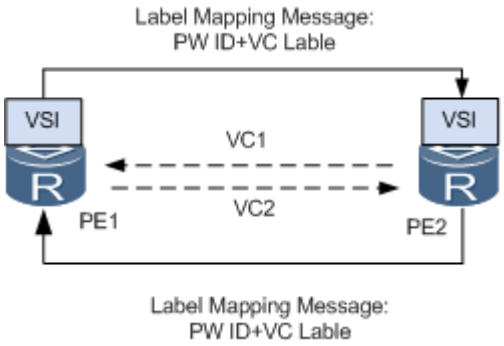
 Copyright © Huawei Technologies Co., Ltd.

from neighboring LSRs that are not at the next hop to re-establish an LSP. This mode requires more memory and label space than the conservative mode.

Implementation Process

- Figure 9-35 shows the process of establishing a PW using LDP signaling.

  Figure 9-35 Establishing a PW using LDP signaling

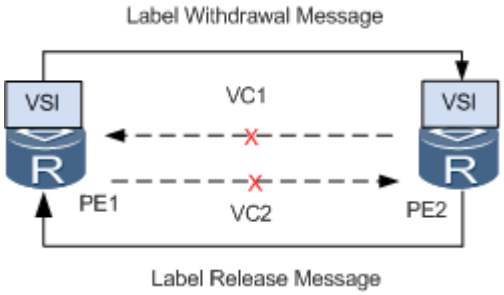
  

  a. After PE1 is associated with a VSI, and PE2 is configured as a peer of PE1, PE1 sends a Label Mapping message to PE2 in DU mode if an LDP session already exists between PE1 and PE2. The Label Mapping message carries information required to establish a PW, such as the PW ID, VC label, and interface parameters.

  b. Upon receipt of the message, PE2 checks whether itself has been associated with the VSI. If PE2 has been associated with the VSI and PW parameters on PE1 and PE2 are consistent, PE1 and PE2 belong to the same VSI. In this case, PE2 establishes a unidirectional VC named VC1 immediately after PE2 receives the Label Mapping message. Meanwhile, PE2 sends a Label Mapping message to PE1. After receiving the message, PE1 takes a similar sequence of actions to PE2 and establishes VC2.

- Figure 9-36 shows the process of tearing down a PW using LDP signaling.

  Figure 9-36 Tearing down a PW using LDP signaling

  a. After the peer configuration about PE2 is deleted from PE1, PE1 sends a Label Withdrawal message to PE2. After receiving the Label Withdrawal

 Copyright © Huawei Technologies Co., Ltd.

message, PE2 withdraws its local VC label, tears down VC1, and sends a Label Release message to PE1.

b. After receiving the Label Release message, PE1 withdraws its local VC label and tears down VC2.

Derivative Functions

Compatibility and switchover between PWE3 VPLS and LDP VPLS

The differences between PWE3 and LDP VPLS are as follows:

- In PWE3 mode, the Label Mapping Message is sent as long as the PW is configured. In LDP VPLS mode, the message is sent only when the local up conditions are met.
- In LDP VPLS, the Label Mapping Message packet does not contain the PW Status. The Label Mapping Message and Label Withdrawal Message are used to notify the local device of the Up and Down states.
- LDP VPLS does not support notification messages to notify status constraints.

MAC Withdrawal

- After receiving a MAC-Withdraw message that carries the NULL MAC TLV, the remote PE clears all MAC address entries in the VSI by default. You can configure a PE to delete MAC address entries in standard mode defined in RFC 4762. In standard mode, only MAC address entries for those ports that are not used by the corresponding PW are deleted.
- After receiving a MAC-Withdraw message that carries the PE-ID TLV, the remote PE clears the MAC address entry for the corresponding PW.

Ignorance of the AC Status by a VSI

Before the replacement of CEs, you can configure VSIs on UPEs to temporarily ignore the AC interface status check. Then, check whether VSIs on UPEs can work properly after new CEs are deployed. A VSI can be Up only if at least one AC interface and one PW are Up. After you configure a VSI to ignore the AC interface status check, the VSI remains Up as long as one PW is Up, regardless of whether the AC interface status is Up or Down.

Receiving of Group Messages by PWs

The IETF defines the usage scenario of this function. If multiple PWs, belonging to the same group and having the same status, are configured on a physical interface, Group messages can be used to notify PWs of the interface status change when the physical interface goes Up or Down, reducing the number of Notification messages required.

NOTE

At present, the MA5800 can only receive group messages and cannot send group messages.

PW Reliability

LDP VPLS ensures PW reliability by manual configuration. When the primary PW fails, traffic from the primary PW switch to the secondary PW; When primary PW recovers, traffic can be immediate or delayed switch back to the primary PW.

 Copyright © Huawei Technologies Co., Ltd.

Usage Scenario

The LDP mode applies to VPLS networks that do not have many sites, do not span multiple ASs, or with PEs that do not run BGP.

Benefits

LDP VPLS brings the following benefits:

- Easy configuration
- Label resource saving

9.3.3.4 VPLS PW Redundancy

Implementation

To ensure the same forwarding capability, the PW redundancy protection mechanism to be used must allow the configuration of a single PW in a PW group to be an active PW and the remaining to be standby PWs, which requires corresponding signaling control.

RFC 4447 (Pseudowire Setup and Maintenance Using the Label Distribution Protocol [LDP]) specifies the PW Status TLV to transmit the PW forwarding status. The PW Status TLV is transported to the remote PW peer using a Label Mapping or LDP Notification message. The PW Status TLV is a 32-bit status code field. Each bit in the status code field can be set individually to indicate more than one failure. PW redundancy introduces a new PW status code 0x00000020. When the code is set, it indicates "PW forwarding standby".

Forwarding priorities (Primary or Secondary) must be configured for PWs that back up each other. The highest priority PW will be selected as the primary PW to forward traffic. The remaining PWs will be in the Secondary state to protect the primary PW.

NOTE

Currently, only one secondary PW can be configured for a primary PW.

The forwarding status of a PW determines whether the PW is used to forward traffic. The PW forwarding statuses depend on:

- Local and remote PW signaling statuses: A PE monitors the local signaling status and uses PW redundancy signaling to obtain remote signaling status from a remote PE.
- PW redundancy mode: Master/Slave or Independent mode is specified on PE1.
- PW forwarding priorities: PW forwarding priorities (Primary or Secondary) are specified on PE1.

Figure 9-37 shows that VPLS PW redundancy is configured on PE1. In normal cases, all local and remote PW signaling statuses on PE1 are Up. PEs at the two ends of a PW in different VPLS PW redundancy modes use different methods to select the same PW for transmitting user packets.

- In Master/Slave mode, PE1 determines local PW forwarding statuses based on preset forwarding priorities and informs PE2 and PE5 of the PW forwarding

 Copyright © Huawei Technologies Co., Ltd.

statuses; PE2 and PE5 determine their PW forwarding statuses based on the received PW primary and secondary statuses.

- In Independent mode, PE1 determines local PW forwarding statuses based on the forwarding statuses learned from PE2 and PE5; PE2 and PE5 determine their PW primary and secondary statuses based on signaling, which can be enhanced trunk (E-Trunk), enhanced automatic protection switching (E-APS), or Virtual Router Redundancy Protocol (VRRP) signaling, and notifies PE1 of the forwarding statuses.

In both Master/Slave and Independent modes, if a primary PW is faulty, it becomes inactive and its secondary PW becomes active. PW-side faults do not affect the AC status. If AC-side faults occur (for example, a PE or AC link is faulty), the PW primary and secondary statuses in Independent mode will change because the statuses are determined by the master and backup statuses of the dual-homing devices; the PW primary and secondary statuses in Master/Slave mode will not change because they are determined by PW side.

NOTE

VPLS PW redundancy is similar to VPWS PW redundancy, with the exception that a virtual switch instance (VSI) has multiple PWs to different PEs. These PWs form various PW groups. PW switching in one group does not affect other PW groups.

Derivative Function

In addition to protection against network faults in real time, VPLS PW redundancy allows users to manually switch traffic between PWs in a group during network operation and maintenance. For example, if a device providing a primary PW needs to be maintained, a user can switch traffic to the secondary PW and switch it back to the primary PW after the maintenance.

NOTE

The interval between a switchover and a switchback must be at least 15s.

Usage Scenarios

VPLS PW redundancy can be used on hierarchical virtual private LAN service (HVPLS) networks and VPLS and virtual leased line (VLL) interconnected networks. These two types of networks can bear any services, but when newly planned or deployed, these networks are suggested to carry different services based on their networking characteristics.

- HVPLS networks are suitable for bearing multicast services, such as Internet Protocol television (IPTV) services, because HVPLS networks can save VPLS core network bandwidth. For details, see **9.3.4.3 VPLS PW Redundancy for Protecting Multicast Services**.
- VPLS and VLL interconnected networks are suitable for bearing unicast services, such as high-speed internet (HSI) and voice over IP (VoIP) services, because VLL PEs do not need to learn user MAC addresses. For details, see **9.3.4.4 VPLS PW Redundancy for Protecting Unicast Services**.

VPLS PW redundancy can also be used to improve reliability of existing networks. On the VPLS network in **Figure 9-37**, CE1 communicates with CE2, CE3, and CE4 through PWs between one VSI on PE1 and PE2, PE3, and PE4.

 Copyright © Huawei Technologies Co., Ltd.

As services develop, services between CE1 and CE2, and between CE1 and CE3 require high reliability. Services between CE1 and CE4 do not require high reliability.

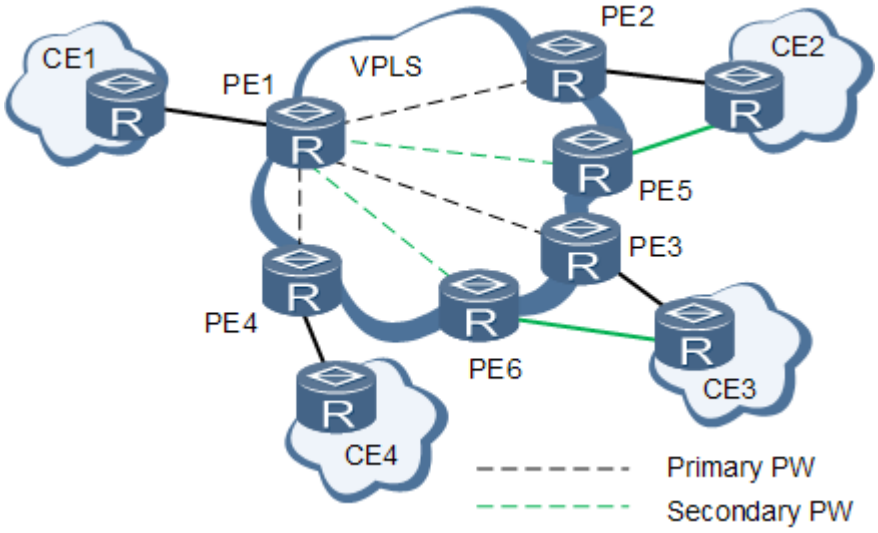
To meet the reliability requirements, PE5 and PE6 are deployed on the VPLS network to provide VPLS PW redundancy protection for PE2 and PE3, respectively. In addition, multiple PW groups to peer PEs are configured in one VSI on PE1. Links between CE1 and CE4 remain unchanged.

VPLS PW redundancy protects services against failures on the network side, AC side, or PEs without affecting existing services, improving network reliability.

NOTE

VPLS PW redundancy can be provided for the desired services without affecting services on other PWs, which reduces costs and maximizes profits.

Figure 9-37 VPLS PW redundancy networking



9.3.4 VPLS PW Redundancy Applications

9.3.4.1 Application of VPLS Individual Access

Service Overview

The traffic of individual services such as high speed internet (HSI), voice over IP (VoIP) and broadband TV (BTV) are carried by the carrier's metropolitan area network (MAN).

The traditional bearing technologies such as the asynchronous transfer mode (ATM) and frame relay (FR) have some defects such as high cost for network construction, slow speed and complicated deployment. Moreover, the traditional bearing technologies only support the point-to-point (P2P) interconnection for users. With the development of IP technology, the Ethernet-based virtual private LAN service (VPLS) technology supports transparent transmission of the above-mentioned individual services and achieves the point-to-multipoint (P2MP)

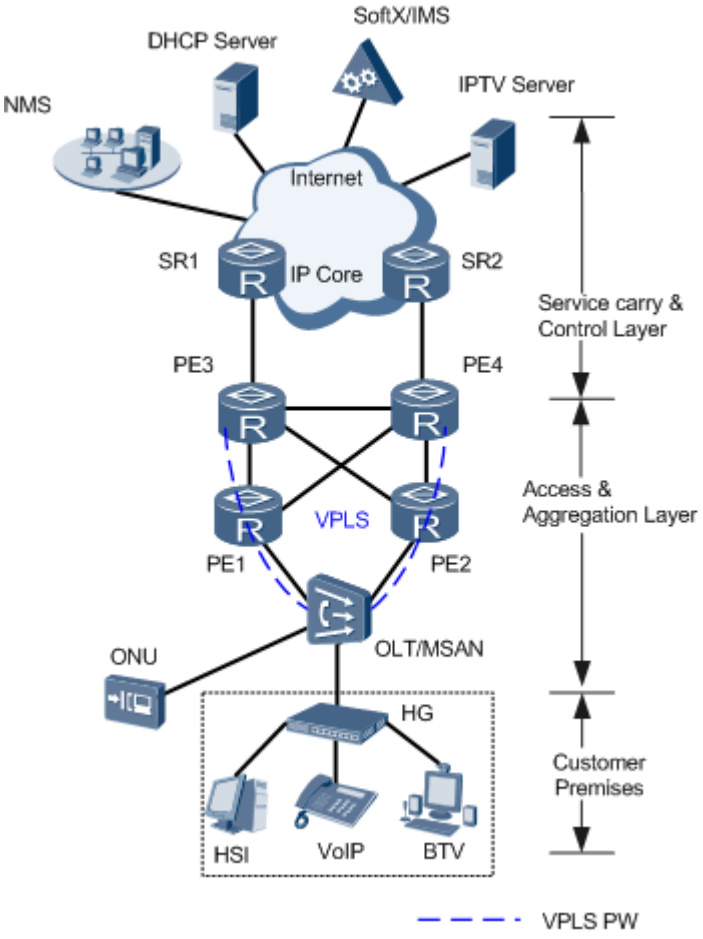
 Copyright © Huawei Technologies Co., Ltd.

interconnection for users. In addition, the Ethernet-based VPLS has many advantages, such as low cost for network construction, high speed and simple deployment. Therefore, the VPLS technology is widely used in the current MAN to transmit the user traffic.

Example Network

Figure 9-38 shows the VPLS individual access service.

Figure 9-38 Example network of VPLS individual access



The HSI service is used as an example in the example network.

- The MSAN/OLT is dual-homed to two AGS devices through the VPLS.
- The user HSI access service is provided through the PPPoE dialup and maps to the VPLS domain through a VLAN in upstream direction.
- PADI packets initiated from the user side are broadcast in the VPLS domain to which the packets belong. The broadcast packets are received on PE1 and PE2.
- The delay response is used between PE devices to terminate the dialups of some users so that the load sharing can be achieved.
- The split horizon between the VPLS and PW is enabled.

 Copyright © Huawei Technologies Co., Ltd.

9.3.4.2 Application of VPLS Enterprise Access

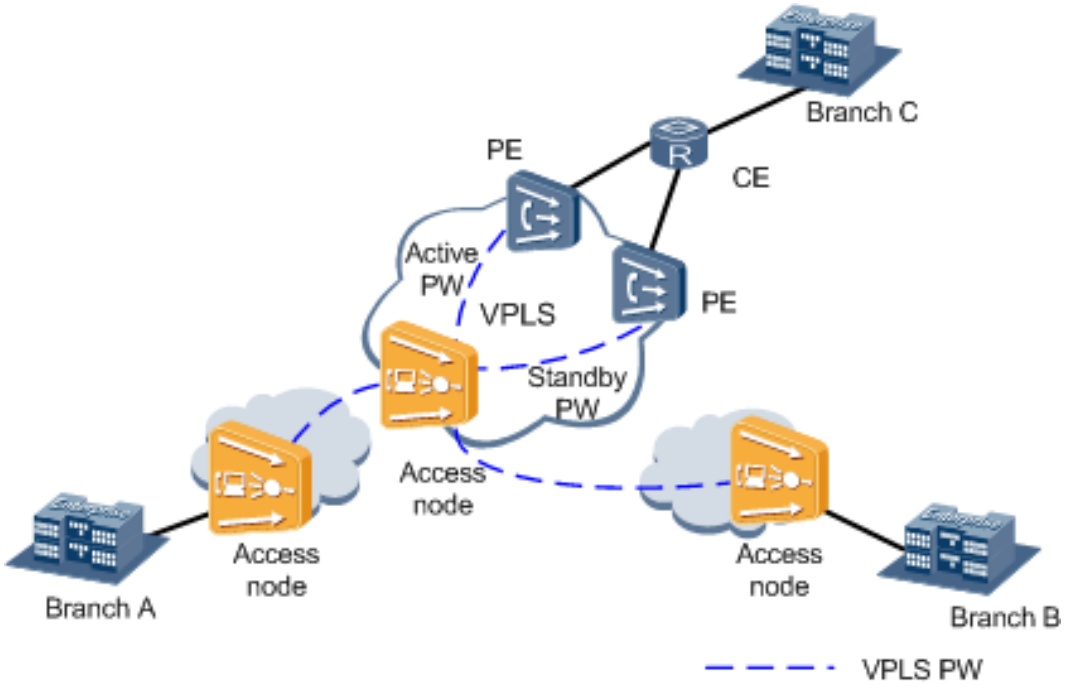
Service Overview

With the business expansion, many enterprises establish branches in different areas and employees are often on business trips. Therefore, some applications (such as the VoIP, instant messages and network conference) are used widely in enterprises. These applications require a network that supports point-to-multipoint (P2MP) services. In addition, the network reliability must be ensured and a transparent and secure data channel is required for multi-point transmission because of the privacy of the enterprise business data. The VPLS technology is suitable to be deployed in this scenario.

Example Network

Figure 9-39 shows the example network of VPLS enterprise access.

Figure 9-39 Example network of VPLS enterprise access



- The virtual private network (VPN) between different branches is achieved by deploying the VPLS.
- The pseudo wire (PW) redundancy is used to protect the important branches (such as branch C in the figure).
- An OLT/MSAN, functioning as the main node, implements the Layer 2 label switching, and other branches are connected to the VPLS network through backup PWs.
- The split horizon between the VPLS and PW is canceled.
- The basic Layer 2 forwarding mechanism in this scenario is consistent with that in the VPLS individual access scenario except that the split horizon needs

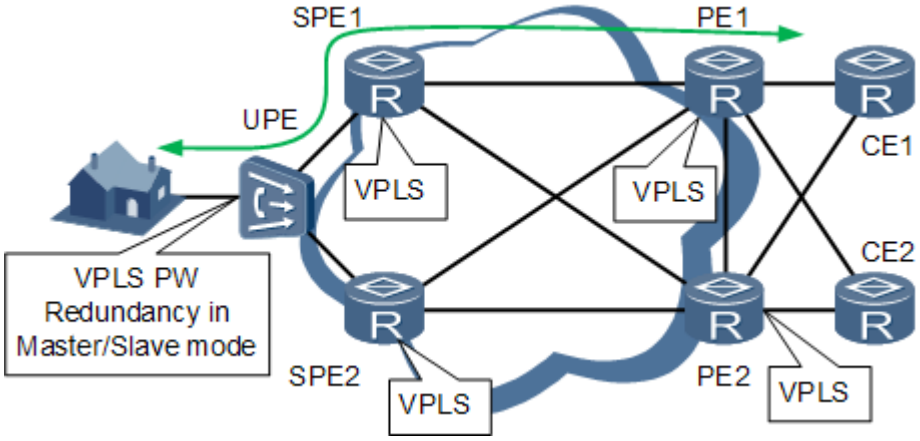
 Copyright © Huawei Technologies Co., Ltd.

to be canceled and the PW protection needs to be supported for Layer 2 forwarding in this scenario.

9.3.4.3 VPLS PW Redundancy for Protecting Multicast Services

Figure 9-40 illustrates an application of VPLS PW redundancy for protecting multicast services, such as Internet Protocol television (IPTV) services, on a hierarchical virtual private LAN service (HVPLS) network. VPLS PW redundancy is used to ensure network reliability.

Figure 9-40 VPLS PW redundancy for protecting multicast services



Multicast sources CE1 and CE2 are each dual-homed to provider edge 1 (PE1) and PE2 through enhanced trunks (E-Trunks); common PWs connect PEs and superstratum PEs (SPEs). A gateway user-end provider edge (UPE) connects the user end to SPE1/SPE2. The link between the UPE and SPE1 and the link between the UPE and SPE2 back up each other.

For this networking, the UPE must use PW redundancy in Master/Slave mode because SPE1 and SPE2 do not exchange signaling for determining which one is the master SPE. Upon detecting that the primary PW fails, the UPE rapidly switches traffic to the secondary PW, instructs SPE2 to work as the primary SPE, and sends MAC Withdraw messages to SPE2 instructing SPE2 to delete the MAC addresses learned from SPE1. SPE2 transmits the MAC Withdraw messages to PE1 and PE2, instructing PE1 and PE2 to clear the MAC addresses learned from SPE1. After deleting the MAC addresses learned from SPE1, PE1 will relearn MAC addresses through multicast packets upon receiving traffic from CE1 and CE2 and switch received traffic to the secondary link.

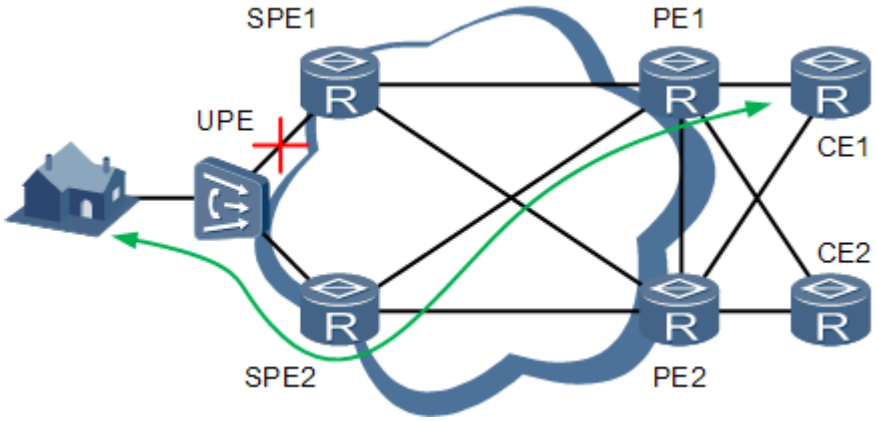
Figure 9-40 shows service traffic when no fault occurs. The following describes how VPLS PW redundancy protects traffic after faults occur.

Primary PW Failure Between the UPE and SPE1

Figure 9-41 shows how traffic is switched if the primary PW between the UPE and SPE1 fails.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-41 VPLS PW redundancy protecting services against a failure in the primary PW between the UPE and SPE1



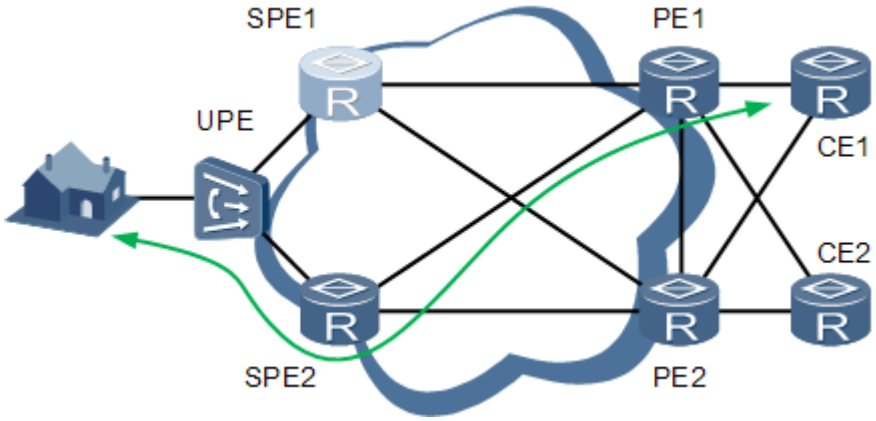
Label switched path (LSP) down events or BFD for PW may cause a PW failure. Upon detecting that the primary PW fails, the UPE switches traffic to the secondary PW and sends MAC Withdraw messages in which the PE ID field carries the SPE1 LSR ID to SPE2. SPE2 transparently transmits the MAC Withdraw messages to PE1 and PE2. SPE2, PE1, and PE2 clear the MAC addresses learned from SPE1.

Switchback: After the primary PW recovers, the UPE instructs SPE2 to change its PW status to secondary and SPE1 to change its PW status to primary. The UPE sends MAC Withdraw messages in which the PE ID field carries the SPE2 LSR ID to SPE1. SPE1 transparently transmits the MAC Withdraw messages to PE1 and PE2. SPE1, PE1, and PE2 clear the MAC addresses learned from SPE2. PE1 and PE2 then relearn MAC addresses through multicast packets from the primary PW.

SPE1 Failure

Figure 9-42 shows how traffic is switched if SPE1 fails.

Figure 9-42 VPLS PW redundancy protecting services against an SPE1 failure



After detecting that SPE1 fails, the UPE switches traffic to the secondary PW and sends MAC Withdraw messages in which the PE ID field carries the SPE1 LSR ID to SPE2. SPE2 transparently transmits the MAC Withdraw messages to PE1 and PE2. SPE2, PE1, and PE2 clear the MAC addresses learned from SPE1. Sometimes, PE1

 Copyright © Huawei Technologies Co., Ltd.

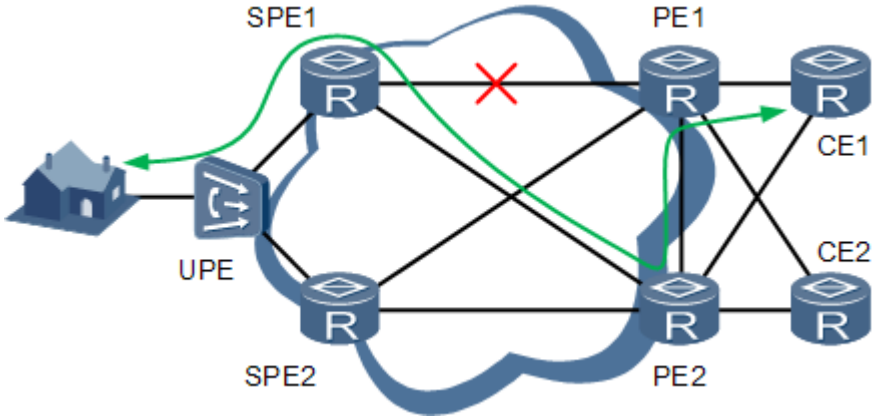
and PE2 detect that the PW passing through SPE1 is faulty before receiving the MAC Withdraw messages and directly clear the MAC addresses learned from SPE1.

Switchback: After the primary PW recovers, the UPE instructs the PW passing through SPE2 to work as a secondary PW and the PW passing through SPE1 to work as the primary PW. The UPE sends MAC Withdraw messages in which the PE ID field carries the SPE2 LSR ID to SPE1. SPE1 transparently transmits the MAC Withdraw messages to PE1 and PE2. SPE1, PE1, and PE2 clear the MAC addresses learned from SPE2. PE1 and PE2 then relearn MAC addresses through multicast packets from the primary PW.

Link Failure Between SPE1 and PE1

Figure 9-43 shows how traffic is switched if the link between SPE1 and PE1 fails.

Figure 9-43 VPLS PW redundancy protecting services against a link failure between SPE1 and PE1



If deployed between SPE1 and PE1, Label Distribution Protocol (LDP) fast reroute (FRR) ensures the availability of traffic between SPE1 and PE1. If LDP FRR is not deployed, LDP LSP ensures the availability of traffic between SPE1 and PE1.

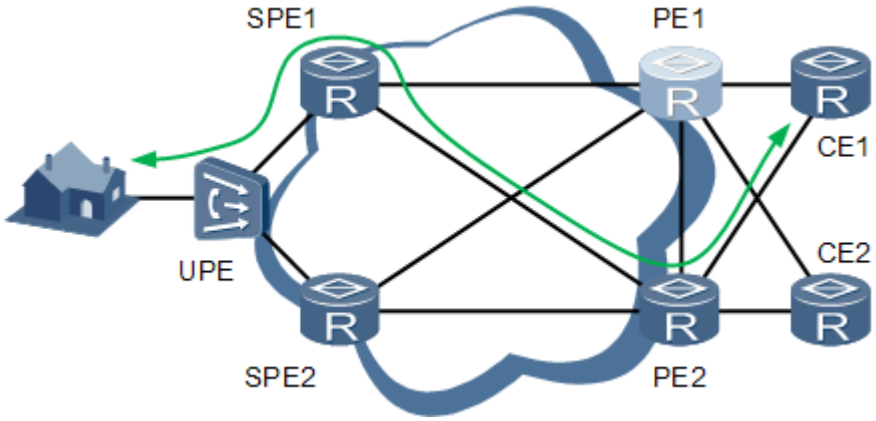
Switchback: Traffic will not be switched back to the primary PW. After LDP LSP convergence, the primary PW is carried by a new LSP.

PE1 Failure

Figure 9-44 shows how traffic is switched if PE1 fails.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-44 VPLS PW redundancy protecting services against a PE1 failure



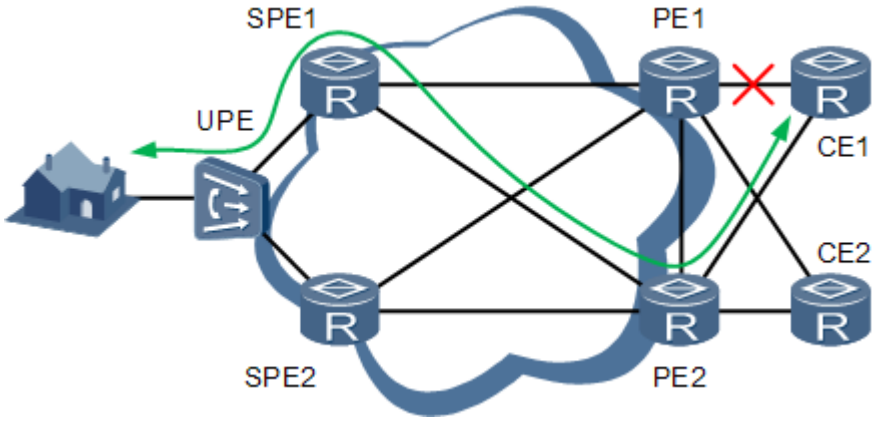
CE1 and CE2 are each dual-homed to PE1 and PE2 through E-Trunk. If PE1 fails, the E-Trunk primary and secondary status changes. PE2 detects E-Trunk status changes and sends MAC Withdraw messages to SPE1 and SPE2, instructing SPE1 and SPE2 to clear MAC addresses in the associated VSI. Sometimes, SPE1 and SPE2 detect that the PW passing through PE1 is faulty before receiving the MAC Withdraw messages and directly clear MAC addresses associated with the PW.

Switchback: If PE1 recovers, traffic switches back to PE1 after a default E-Trunk switchback delay. Upon detecting E-Trunk changes, PE1 and PE2 send MAC Withdraw messages to SPE1 and SPE2, instructing SPE1 and SPE2 to clear MAC addresses learned from PE2.

Primary AC Link Failure

Figure 9-45 shows how traffic is switched if the link between CE1 and PE1 fails.

Figure 9-45 VPLS PW redundancy protecting services against a link failure between CE1 and PE1



CE1 and CE2 are each dual-homed to PE1 and PE2 through E-Trunk. After the link between CE1 and PE1 fails, the E-Trunk primary and secondary status changes. Upon detecting E-Trunk status changes, PE1 and PE2 send MAC Withdraw messages to SPE1, instructing SPE1 to clear all MAC addresses in the VSI.

Switchback: After the link between CE1 and PE1 recovers, the E-Trunk primary and secondary status changes back. Upon detecting E-Trunk changes, PE1 and PE2

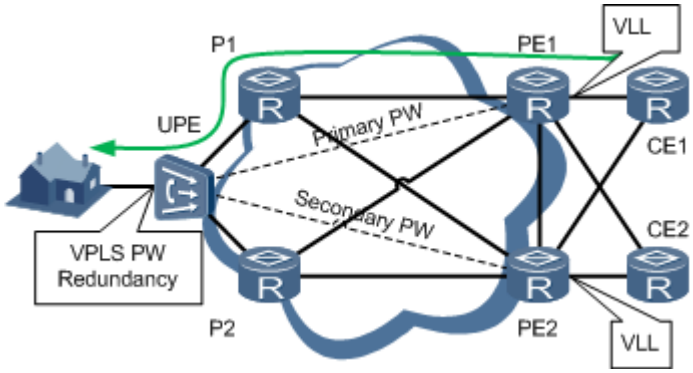
 Copyright © Huawei Technologies Co., Ltd.

send MAC Withdraw messages to SPE1, instructing SPE1 to clear all MAC addresses.

9.3.4.4 VPLS PW Redundancy for Protecting Unicast Services

Figure 9-46 illustrates an application of VPLS PW redundancy for protecting unicast services, such as high-speed internet (HSI) or voice over IP (VoIP) services, on a virtual private LAN service (VPLS) and virtual leased line (VLL) interconnected network. VPLS PW redundancy is used to ensure network reliability.

Figure 9-46 VPLS PW redundancy for protecting unicast services



Authentication servers CE1 and CE2 are each dual-homed to PE1 and PE2 through enhanced trunks (E-Trunks). A UPE connects the user end and PE1/PE2. The link between the UPE and PE1 and the link between the UPE and PE2 back up each other.

For this networking, PE1 and PE2 can determine their master and backup statuses through E-Trunk negotiation. Therefore, the UPE can use PW redundancy in Independent mode to determine the PW primary and secondary statuses based on the status of PE1 and PE2. Upon detecting that the primary PW fails, the UPE rapidly switches traffic to the secondary PW and instructs PE2 to work as the master PE. After the E-Trunk detects that the primary PW fails, it switches traffic to the secondary AC link.

Figure 9-46 shows service traffic when no fault occurs. The following describes how VPLS PW redundancy protects services after a fault occurs.

NOTE

When a fault occurs, VPLS PW redundancy protects services differently in Master/Slave and Independent modes. The following describes the difference in service protection provided by VPLS PW redundancy in Master/Slave and Independent modes. If no description is provided, no difference exists.

Network Side Failure

Label switched path (LSP) or traffic engineering (TE) tunnels and PW redundancy can protect services against faults on the UPE network side.

 Copyright © Huawei Technologies Co., Ltd.

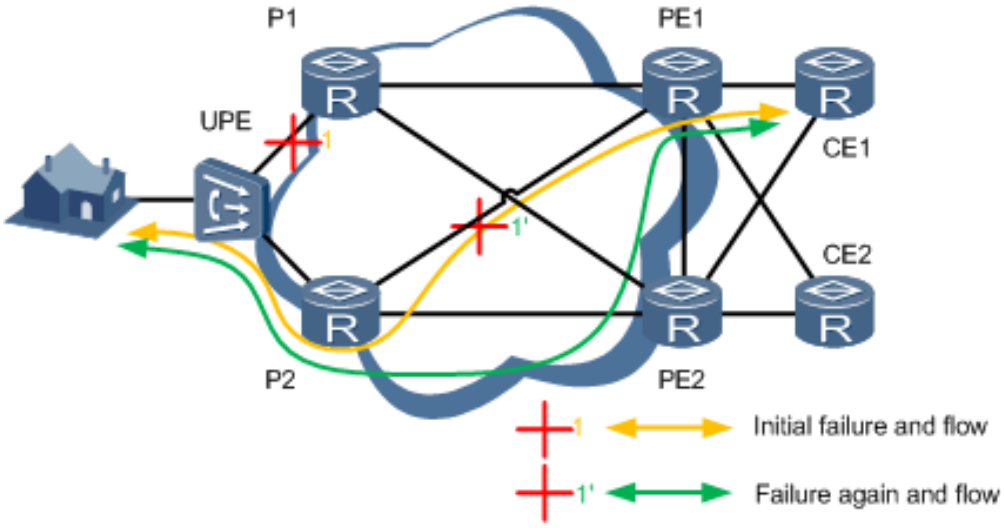
NOTE

If a network-side fault occurs, LSP or TE tunnels detect the fault and switch traffic to other tunnels. If tunnel protection is unavailable or fails, PW redundancy is required to protect traffic. A bypass PW needs to be configured between PE1 and PE2 for PW redundancy.

- LSP or TE tunnel protection: After a network-side fault occurs, routes and LSPs are recomputed. The LSP where the primary PW is located is converged to a new path. **Figure 9-47** shows how traffic is switched.

  After the fault is rectified, LSPs re-converge and the primary PW is carried by a new LSP.

  Figure 9-47 Protecting services against an LSP failure (bypass PW not configured)



- PW redundancy: If LSP or TE tunnel switching fails, traffic is switched to the secondary PW. **Figure 9-48** shows how traffic is switched.

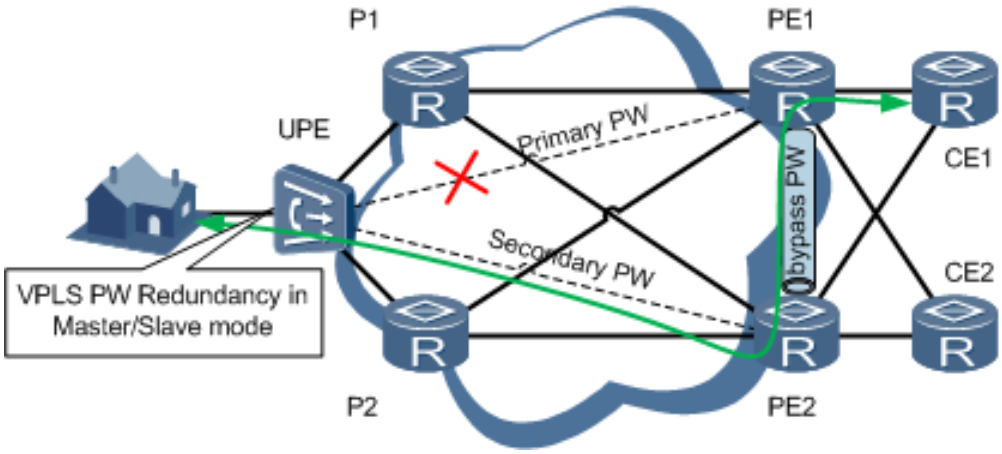
  After the fault is rectified, traffic will be switched back based on preset switchback policies.

  NOTE

  Bypass PWs are required for PW redundancy to transmit traffic between PE1 and PE2.

 Copyright © Huawei Technologies Co., Ltd.

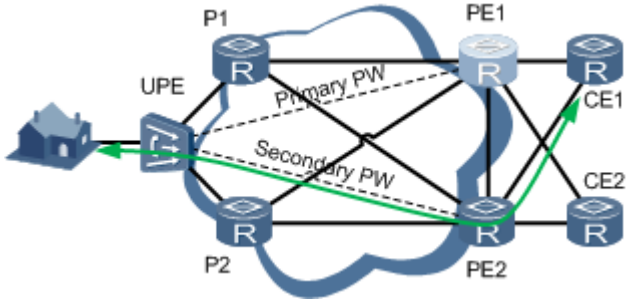
Figure 9-48 Protecting services against a primary PW failure (bypass PW configured)



PE1 Failure

Figure 9-49 shows how traffic is switched if PE1 fails.

Figure 9-49 VPLS PW redundancy protecting services against a PE1 failure



PE2 becomes the master and PE1 becomes the backup after E-Trunk negotiation. The UPE is informed of the switchover. Upon detecting that the primary PW fails, the UPE clears MAC addresses learned from the primary PW and switches traffic to the secondary PW.

Switchback: After the PE1 fault is rectified, PE1 becomes the master through E-Trunk negotiation. Upon detecting PE1 and PE2 status changes, the UPE clears MAC addresses learned from PE2 and relearns MAC addresses through multicast packets.

Primary AC Link Failure

Figure 9-50 shows how traffic is switched if the link between PE1 and CE1 fails.

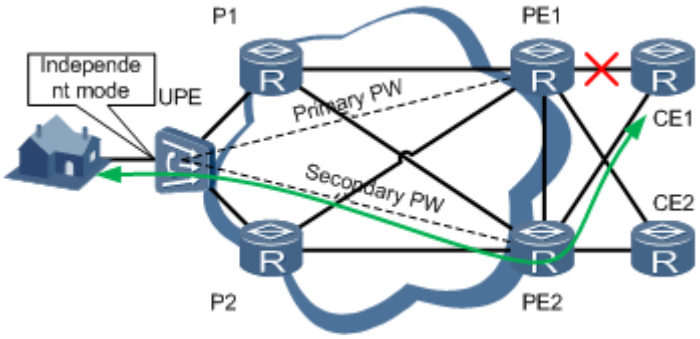
- In Independent mode

  After the primary AC link between CE1 and PE1 fails, PE2 works as the master after E-Trunk negotiation. The UPE is informed of the switchover. The UPE detects that the primary AC link fails and switches traffic to the secondary PW.

 Copyright © Huawei Technologies Co., Ltd.

After the link between CE1 and PE1 recovers, PE1 becomes the master after E-Trunk negotiation. Upon detecting PE1 and PE2 status changes, the UPE clears MAC addresses learned from PE2 and relearns MAC addresses through multicast packets.

Figure 9-50 VPLS PW redundancy in Independent mode protecting services against a failure in the primary AC link

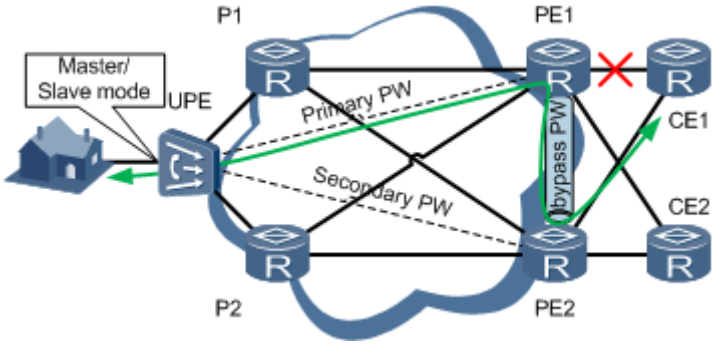


- In Master/Slave mode

  After the primary AC link between CE1 and PE1 fails, PE2 works as the master after E-Trunk negotiation. The PW status on the UPE remains unchanged.

  After the link between CE1 and PE1 recovers, PE1 becomes the master after E-Trunk negotiation. Upon detecting PE1 and PE2 status changes, the UPE clears MAC addresses learned from PE2 and relearns MAC addresses through multicast packets.

  Figure 9-51 VPLS PW redundancy in Master/Slave mode protecting services against a failure in the primary AC link



9.3.5 Configuring VPLS MP2MP Intercommunication

VPLS can implement the multipoint-to-multipoint (MP2MP) VPN networking; therefore, by using the VPLS technology, service providers (SPs) can provide the Ethernet-based multipoint services through MPLS backbone networks.

Application Context

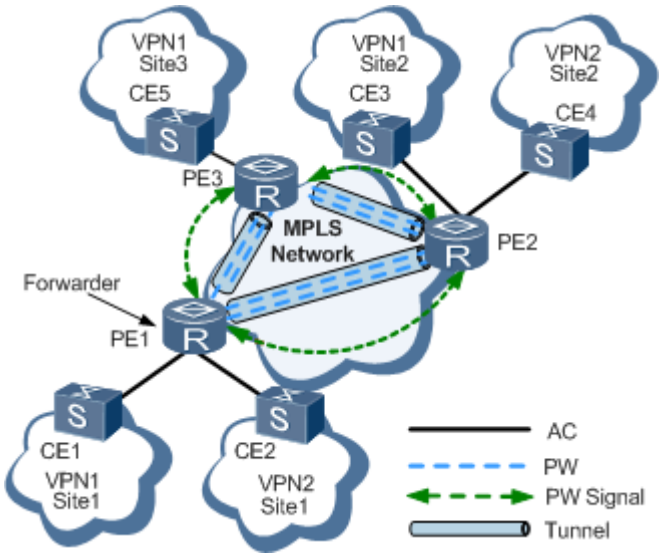
A lot of private line services in carriers' network use the virtual private network (VPN) virtual private wire service (VPWS) technology, which can provide point-to-point (P2P) communication services on Layer 2 or Layer 3 network. With the development of Ethernet and MPLS technologies, carriers hope to provide not only P2P services on the private network, but Ethernet-like point-to-multipoint (P2MP)

 Copyright © Huawei Technologies Co., Ltd.

services on the metropolitan area network (MAN) and wide area network (WAN). By deploying virtual private LAN service (VPLS) technology on the provider edge (PE), carriers can provide Ethernet-based MP2MP services for users through MPLS backbone networks, achieving the local area network (LAN) simulation.

Figure 9-52 shows the basic VPLS transmission process. Full-meshed PWs are created through signaling transmission by PE routers. Transmission of packets between CEs relies on VSIs configured on PEs, and PWs established between the VSIs.

Figure 9-52 Basic VPLS transmission process



Prerequisite

1. The IP address of the loopback interface must be configured.
2. The LSR ID must be configured.
3. The VLAN for MPLS label forwarding must be created.
4. The global MPLS, VLAN MPLS, and VLAN interface MPLS must be enabled.
5. MPLS L2VPN must be enabled.
6. A static or dynamic route must be successfully configured on each device in the network (so that LSRs can reach each other through the IP route).
7. The global LDP function is enabled and remote LDP sessions are configured.
8. The vlan-based traffic stream must be created.

Data preparation

Before configuring the VPLS P2MP intercommunication services, plan the data items as listed in **Table 9-7**.

 Copyright © Huawei Technologies Co., Ltd.

Table 9-7 Plan of VPLS P2MP intercommunication service data items

Item	Data	Remarks
MA5800	VSI	VLANs are mapped to the VPLS domain after a VSI is bound to the VLAN and PW. VLAN mapping allows service packets to be broadcast in the VPLS domain.
	VPLS PW	-

Procedure

Step 1 Configure a VSI.

VSIs are the core of VPLS services. With VSIs, actual links carrying VPLS services can be mapped into PWs.

1. In global config mode, run the **vsi** command to create a VSI and enter the VSI mode.
2. Run the **pwsignal ldp** command to configure the signaling type for VSI as LDP, or run the **bgp-ad** command to configure the signaling type for VSI as Bgp-ad.
3. Run the **vsi-id** command to configure the VSI ID.

   Once the VSI ID is successfully set, it cannot be changed or deleted. If you need to change it, delete the VSI.

Step 2 (Optional) Configure VSI attributes.

In VSI mode, configure VSI basic attributes based on actual requirements. VSI basic attributes include the VSI description information, encapsulation type, control words, maximum transmission unit (MTU), and traffic suppression.

- Run the **description** command to configure the description of a VSI.
- Run the **encapsulation** command to configure the encapsulation type of a VSI.
- Run the **control-word** command to enable the control word of a VSI. After the control word is enabled, control information will be added to packets.

  NOTE

  If you use the **control-word** command in VSI mode and the **control-word** command in PW-para-index mode to configure the control word concurrently, the one set by the **control-word** command in PW-para-index mode takes effect.
- Run the **mtu** command to set the MTU of a VSI.
- Run the **traffic-suppress** command to set the suppression level of the broadcast, unknown multicast, and unknown unicast traffic for a VSI.

  Before configuring the multicast service carried in VPLS, you must disable the VSI unknown multicast suppression. Otherwise, packet loss will occur in the multicast services.

 Copyright © Huawei Technologies Co., Ltd.

Step 3 Configure LDP PW.

1. In global config mode, run the **pw-para pwindex** *pwindex* command to create a PW and enter the PW-para-index mode.

   For a VPLS PW, you must first create the PW-para-index mode and then perform the PW binding.
2. Run the **service-type vpls** command to configure the service type of a PW as VPLS.
3. Run the **pwid** command to configure the ID of a PW.
4. Run the **peer-address** command to set the IP address of the peer device of a PW.
5. Run the **pw-type ethernet** command to configure the type of a PW as Ethernet.

   When the service type is VPLS, you can set the PW type only to **Ethernet**. The PW type must be identical to the VSI encapsulation type.
6. (Optional) Run the **control-word** command to enable the control word of a PW.
7. (Optional) Run the **dyn-receive-label** command to specify the incoming label of a dynamic PW.
8. In VSI mode, run the **vsi-pw-binding** command to bind the VSI to the PW to create a VPLS PW service.

Step 4 Or configure Bgp-ad PW.

1. Run the **bgp-ad** command in VSI mode to specify the current VSI PW setup mode to automatic discovery or automatic deployment and enter the VSI BGP AD mode.
2. Run the **vpls-id** command to allocate the BGP AD mode VSI identifier.
3. Run the **vpn-target** command to associate the current VSI with one or more VPN targets. The Import VPN Target of the local PE must be the same as Export VPN Target of the peer PE. PWs can only be set up between the BGP AD VSI instances.

Step 5 In VSI mode, run the **vsi-ac-binding vlan** command to bind a VLAN to the VSI.

After the above configurations are complete, VLAN service packets can be forwarded within a VSI.

----End

Example

Assume that VLAN 100 is used for MPLS forwarding, a VSI and a PW are created, and the PW and VLAN 100 are bound to the VSI respectively. To configure VSI and PW parameters as follows:

- To set the VSI ID to **1**, the VSI name to **hsi**, and the signaling mode to **LDP**, and retain the default values for other parameters, do as follows:
- To set the PW index to **1**, the service type to **VPLS**, the PW ID to **1**, the IP address of the peer device to **10.1.1.1**, the encapsulation type to **Ethernet tagged**, and the dynamic PW incoming label to **10240**, do as follows:

```
huawei(config)#vsi hsi
huawei(config-vsi-hsi)#pwsignal ldp
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-vsi-hsi-ldp)#vsi-id 1
huawei(config-vsi-hsi-ldp)#quit
huawei(config-vsi-hsi)#quit
huawei(config)#pw-para pwindex 1
huawei(config-pw-para-index-1)#service-type vpls
huawei(config-pw-para-index-1)#pwid 1
huawei(config-pw-para-index-1)#peer-address 10.1.1.1
huawei(config-pw-para-index-1)#pw-type ethernet tagged
huawei(config-pw-para-index-1)#dyn-receive-label 10240
huawei(config-pw-para-index-1)#quit
huawei(config)#vsi hsi
huawei(config-vsi-hsi)#pwsignal ldp
huawei(config-vsi-hsi-ldp)#vsi-pw-binding pwindex 1
huawei(config-vsi-hsi-ldp)#quit
huawei(config-vsi-hsi)#vsi-ac-binding vlan 100
```

9.3.6 Configuration Example of VPLS

Virtual private LAN service (VPLS) enables geographically isolated users (individuals or branch offices of an enterprise) to establish point-to-multipoint connections between each other using Ethernet links, achieving fast and flexible service deployment.

Context

As a combination of Ethernet and Multiprotocol Label Switching (MPLS) technologies, VPLS emulates all functions of the traditional local area network (LAN), with a purpose to connect multiple geographically isolated LANs that consist of Ethernet using the IP or MPLS network provided by carriers and make the LANs work as a LAN.

Currently, the MA5800 only supports Label Distribution Protocol (LDP)-based VPLS.

Table 9-8 provides basic concepts of VPLS.

Table 9-8 Basic concepts of VPLS

Concept	Description
VSI	Virtual switch instance. This concept corresponds to virtual local area network (VLAN) of Ethernet switch. Each VSI provides independent VPLS service. VSI supports the Ethernet bridge function and can terminate PW.

 Copyright © Huawei Technologies Co., Ltd.

Concept	Description
PW	Pseudo wire. PW is a virtual connection between two PEs and transmits frames between the PEs. PEs use signaling to establish PWs and maintain PW status. On a VPLS network, PE routers transmit signaling to each other to establish PWs of full interconnection. Signaling exchange modes are as follows: • Martini mode: In this mode, signaling is exchanged using the LDP protocol. This mode does not support the PE automatic discovery function. PEs need to be configured manually. With this mode, the networking is simple, and low requirements are imposed on PEs which do not need to cross domains. • Kompella mode: Signaling is exchanged using the Border Gateway Protocol (BGP). The MA5800 does not support this mode.
AC	Attachment circuit. AC attaches a CE to a PE. An AC can be a physical or logical link. It transmits frames between the CE and PE.
Split horizon	Split horizon is a technology that prevents route loops and speeds up route convergence. In a VPLS network, full mesh and split horizon are used to prevent loops. Split horizon in VPLS means that the data packets received from the PW at the PSN side are not forwarded to other PWs. Instead, they are forwarded to the private network.

9.3.6.1 Configuration Example of the VPLS Internet Access Service

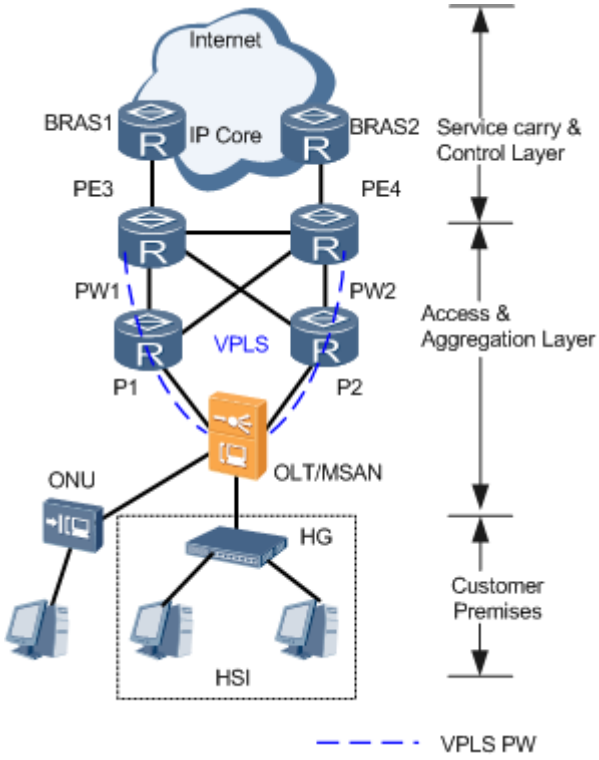
This topic describes how to configure the Internet access service for individual users when the VPLS networking is used at the access and aggregation layers.

Application Context

As shown in **Figure 9-53**, the MA5800 is dual homed to two AGS devices (aggregation switches) PE3 and PE4 through VPLS, and the Internet access service is received in PPPoE dialup mode. In the upstream direction, the traffic stream is mapped into the VPLS domain through VLAN. The PPPOE active discovery initiation (PADI) packets initiated by the user are broadcast in the VPLS domain it belongs to, and then the broadcast packets are received by the two AGS devices. The AGS devices terminate packets of some users respectively in delay response mode to achieve load sharing. When a BRAS device is faulty, the user dials up again. Then the VPLS MAC learning and forwarding mechanism automatically selects a new BRAS to provide services.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-53 Networking for the VPLS Internet access service for individual users



Prerequisite

Traffic streams have been configured on the MA5800 for the Internet access service.

NOTE

To configure the Internet access service, you must configure SVLAN 100-based traffic streams on the MA5800 and perform corresponding configurations on the HG. The configurations are the same as those for common Internet access service, which are not described here.

Data Plan

Table 9-9 provides the key data plan for the MA5800.

 Copyright © Huawei Technologies Co., Ltd.

Table 9-9 Key data plan

Configuration Item	Data	Remarks	Requirement on PE3 and PE4
MPLS	• LSR ID: 10.10.10.10 • VLAN: 4001	MPLS must be enabled at three layers. • MPLS must be enabled globally. • MPLS must be enabled for VLAN. • MPLS must be enabled at VLAN interfaces.	The LSR ID must be unique on the entire network and MPLS must be enabled.
LDP	• MPLS LDP is enabled. • Split horizon is enabled.	MPLS LDP must be enabled at three layers. • MPLS LDP must be enabled globally. • MPLS LDP must be enabled at VLAN interfaces.	MPLS LDP is enabled. The remote LDP session to the MA5800 is configured on PE3 and PE4.
Routing protocol	The Open Shortest Path First (OSPF) protocol is used.	Ensure that the Layer 3 interfaces on the MA5800 and those on PE3 and PE4 can ping each other.	Layer 3 interfaces and routes are configured on PE3 and PE4. Ensure that the Layer 3 interfaces and LSR IDs on PE3 and PE4 and those on the MA5800 can ping each other.
VPLS PW	• PW ID: 1 and 2 • Service type: vpls • Encapsulation type: ethernet tagged • The control word is enabled.	On the MA5800, PW1 and PW2 are created for PE3 and PE4 respectively. In this way, packets can be transmitted to BRAS devices over two trails.	The LDP VPLS is supported. VPLS PWs to the MA5800 are configured on PE3 and PE4, and attributes of PWs are consistent with those on the MA5800.

 Copyright © Huawei Technologies Co., Ltd.

Configuration Item	Data	Remarks	Requirement on PE3 and PE4
VSI	• PW1 and PW2 are bound to VSI. • VLAN 100 of the Internet access service is bound to VSI.	VSI binds VLAN and PW to map VLAN to the VPLS domain, so that PADI packets for the Internet access service can be broadcast in the VPLS domain at first, until one of the BRAS responses.	VSI is configured on PE3 and PE4 and the VSI ID must bind the corresponding PW.

Procedure

Step 1 Configure the basic MPLS.

1. Configure a loopback interface.

   Set the ID of the loopback interface to **0** and its IP address to **10.10.10.10/32**.

   ```
   huawei(config)#interface loopback 0
   huawei(config-if-loopback0)#ip address 10.10.10.10 32
   huawei(config-if-loopback0)#quit
   ```

2. Configure the MPLS LSR-ID. Use the IP address of loopback interface 0 as the LSR ID.

   ```
   huawei(config)#mpls lsr-id 10.10.10.10
   ```

3. Enable MPLS globally.

   Trigger LDP by the IP address of the host to set up an LSP.

   ```
   huawei(config)#mpls
   huawei(config-mpls)#lsp-trigger host
   huawei(config-mpls)#quit
   ```

4. Enable the L2VPN function.

   ```
   huawei(config)#mpls l2vpn
   ```

5. Enable the LDP function globally and enable the split horizon policy.

   ```
   huawei(config)#mpls ldp
   huawei(config-mpls-ldp)#outbound peer all split-horizon
   huawei(config-mpls-ldp)#quit
   ```

Step 2 Configure VLAN, and enable MPLS for VLAN and VLAN interfaces.

1. Add VLAN 4001 for forwarding MPLS packets and add upstream port 0/9/0 and 0/9/1 to it.

   ```
   huawei(config)#vlan 4001 smart
   huawei(config)#port vlan 4001 0/9/0
   huawei(config)#port vlan 4001 0/9/1
   ```

2. Enable MPLS for VLAN 4001. Set the IP address of VLAN interface 4001 to 10.50.50.50/24 and enable MPLS LDP for the VLAN interface.

   ```
   huawei(config)#interface vlanif 4001
   huawei(config-if-vlanif4001)#ip address 10.50.50.50 24
   huawei(config-if-vlanif4001)#mpls
   ```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-if-vlanif4001)#mpls ldp
huawei(config-if-vlanif4001)#quit
```

Step 3 Configure routes.

VPLS has no special requirements on routing policy. You can use static route, RIP, or OSPF policy. In the following example, OSPF is used.

Set the OSPF process ID to **100** and OSPF area ID to **1**. In addition, configure the interfaces (VLAN interface and loopback interface) that run OSPF and configure the areas of the interfaces.

```
huawei(config)#ospf 1
huawei(config-ospf-1)#area 100
huawei(config-ospf-1-area-0.0.0.100)#network 10.50.50.0 0.0.0.255
huawei(config-ospf-1-area-0.0.0.100)#network 10.10.10.10 0.0.0.0
huawei(config-ospf-1-area-0.0.0.100)#return
```

Step 4 Configure the remote LDP session.

Configure the remote LDP session from the MA5800 to PE3 (LSR ID: 10.3.3.3) and PE4 (LSR ID: 10.4.4.4) respectively. Name the sessions **to_pe3** and **to_pe4** respectively.

```
huawei(config)#mpls ldp remote-peer to_pe3
huawei(config-mpls-ldp-remote-to_pe3)#remote-ip 10.3.3.3
huawei(config-mpls-ldp-remote-to_pe3)#remote-ip auto-dod-request
huawei(config-mpls-ldp-remote-to_pe3)#quit
huawei(config)#mpls ldp remote-peer to_pe4
huawei(config-mpls-ldp-remote-to_pe4)#remote-ip 10.4.4.4
huawei(config-mpls-ldp-remote-to_pe4)#remote-ip auto-dod-request
huawei(config-mpls-ldp-remote-to_pe4)#quit
```

Step 5 Configure a VSI.

1. Add a VSI.

   Create a VSI named **hsi**. Set the signaling protocol to LDP and VSI ID to 1.

   ```
   huawei(config)#vsi hsi
   huawei(config-vsi-hsi)#pwsignal ldp
   huawei(config-vsi-hsi-ldp)#vsi-id 1
   ```

2. (Optional) Configure the attributes of VSI.

   Configure basic attributes of VSI as required, including the encapsulation type, control word, MTU value, and traffic suppression policy. In the following example, the control word is enabled and default values are used for other parameters.

   ```
   huawei(config-vsi-hsi)#control-word
   ```

Step 6 Configure PWs.

Create two PWs with IDs 1 and 2. Set the service type to **vpls**, the remote IP addresses to **10.3.3.3** and **10.4.4.4** respectively, and encapsulation type to **ethernet tagged**. Enable the control word, and set the receive labels of dynamic PW to **10240** and **10250** respectively.

```
huawei(config)#pw-para pwindex 1
huawei(config-pw-para-index-1)#service-type vpls
huawei(config-pw-para-index-1)#pwid 1
huawei(config-pw-para-index-1)#peer-address 10.3.3.3
huawei(config-pw-para-index-1)#pw-type ethernet tagged
huawei(config-pw-para-index-1)#control-word enable
huawei(config-pw-para-index-1)#dyn-receive-label 10240
huawei(config-pw-para-index-1)#quit
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#pw-para pwindex 2
huawei(config-pw-para-index-2)#service-type vpls
huawei(config-pw-para-index-2)#pwid 2
huawei(config-pw-para-index-2)#peer-address 10.4.4.4
huawei(config-pw-para-index-2)#pw-type ethernet tagged
huawei(config-pw-para-index-2)#control-word enable
huawei(config-pw-para-index-2)#dyn-receive-label 10250
huawei(config-pw-para-index-2)#quit
```

Step 7 Bind PW and VSI.

Dynamically bind PW1 and PW2 to the VSI named **hsi** to establish the VPLS PW service.

```
huawei(config)#vsi hsi
huawei(config-vsi-hsi)#pwsignal ldp
huawei(config-vsi-hsi-ldp)#vsi-pw-binding pwindex 1
huawei(config-vsi-hsi-ldp)#vsi-pw-binding pwindex 2
```

Step 8 Bind AC and VSI.

Bind VLAN 100 to the VSI named hsi, so that the Internet access service packets of VLAN 100 can be forwarded in the VSI.

```
huawei(config-vsi-hsi)#vsi-ac-binding vlan 100
```

----End

Result

1. A user performs PPPoE dialup. A PADI packet is transmitted upstream.
2. The OLT maps the PADI packet to VSI based on SVLAN carried in the packet and broadcasts the packet on two PWs. At the same time, the OLT learns user's MAC address.
3. After receiving the PADI packet, two BRAS devices respond with PPPOE active discovery offer (PADO) packets in random delay mode.
4. After receiving the PADO packets from PWs, the OLT learns the MAC address carried in the packets from PWs and forwards the packets to the user.
5. The user receives two PADO packets at different time and only responds to the first received PADO packet to establish a PPPoE session.
6. The OLT forwards subsequent PPPOE active discovery request (PADR) and PPPOE active discovery session-confirmation (PADS) packets based on learnt MAC addresses.

9.3.6.2 Configuration Example of the VPLS Multicast Service

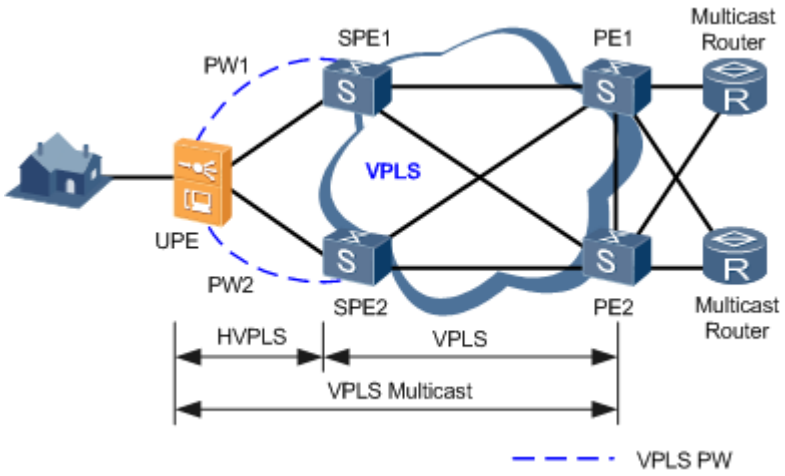
In the networking for multicast services, VPLS deployed on the MA5800 enables the MA5800 to be dual homed to AGS devices (aggregation switches), which ensures network reliability.

Application Context

In the VPLS network, multicast services are still deployed based on multicast VLAN other than based on VSI (VPLS instance). The multicast VLAN is bound to VSI for carrying multicast services over the VPLS network, including upstream IGMP packets and downstream multicast traffic streams.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-54 Networking for the VPLS multicast service



As shown in **Figure 9-54**:

- The MA5800, functioning as a UPE (underlayer PE), is located on the edge network of HVPLS. In the downstream direction, the MA5800 accesses multicast users through multicast VLAN whose IGMP mode is proxy or snooping. In the upstream direction, the MA5800 is dual homed to two AGS devices that serve as SPEs (superstratum PEs) through two PWs. The same VSI is set up on the MA5800 and AGS and the two upstream PWs belong to the VSI.
- Multiple AGS devices form a VPLS core bearer network. The AGS devices learn the VPLS multicast forwarding table and duplicate multicast traffic based on VPLS PW connections. Two edge AGS devices functioning as SPEs in the downstream direction access the UPE on the edge network, and two edge AGS devices functioning as PEs in the upstream direction terminate VPLS and are connected to multicast routers.
- Two multicast routers back up each other and import multicast traffic streams from the multicast source to the VPLS bearer network.

NOTE

Currently, a maximum of two PWs are supported for VPLS multicast services and both two PWs are used for transmitting multicast packets upstream.

Prerequisite

The MA5800 has been configured with multicast VLAN 100-based multicast traffic streams, and corresponding configurations have been performed on the HG. The configurations are similar to those for common multicast services. The difference is that multicast upstream ports do not need to be configured for VPLS multicast services and the default mode is used for the multicast upstream ports.

Data Plan

Table 9-10 provides the key data plan for the MA5800.

 Copyright © Huawei Technologies Co., Ltd.

Table 9-10 Key data plan

Configuration Item	Data	Remarks	Requirement on SPE1 and SPE2
MPLS	• LSR ID: 10.10.10.10 • VLAN: 4001	MPLS must be enabled at three layers. • MPLS must be enabled globally. • MPLS must be enabled for VLAN. • MPLS must be enabled at VLAN interfaces.	The LSR ID must be unique on the entire network and MPLS must be enabled.
LDP	• MPLS LDP is enabled. • Split horizon is enabled.	MPLS LDP must be enabled at two layers. • MPLS LDP must be enabled globally. • MPLS LDP must be enabled at VLAN interfaces.	MPLS LDP is enabled. The remote LDP session to the MA5800 is configured on SPE1 and SPE2.
Routing protocol	The Open Shortest Path First (OSPF) protocol is used.	Ensure that the Layer 3 interfaces on the MA5800 and those on SPE1 and SPE2 can ping each other.	Layer 3 interfaces and routes are configured on SPE1 and SPE2. Ensure that the Layer 3 interfaces and loopback interfaces on SPE1 and SPE2 and those on the MA5800 can ping each other.
VPLS PW	• PW ID: 1 and 2 • Service type: vpls • Encapsulation type: ethernet tagged • The control word is enabled.	L2VPN must be enabled. On the MA5800, PW1 and PW2 are created for SPE1 and SPE2 respectively.	The LDP VPLS is supported. VPLS PWs to the MA5800 are configured on SPE1 and SPE2, and attributes of PWs are consistent with those on the MA5800.

 Copyright © Huawei Technologies Co., Ltd.

Configuration Item	Data	Remarks	Requirement on SPE1 and SPE2
VSI	• PW1 and PW2 are bound to VSI. • Multicast service VLAN 100 is bound to VSI.	Multicast VLAN is bound through VSI to map VLAN to the VPLS domain, so that multicast service packets can be broadcast in the VPLS domain.	VSI is configured on SPE1 and SPE2 and the VSI ID must bind the corresponding PW.

Procedure

Step 1 Configure the basic MPLS.

1. Configure a loopback interface.

   Set the ID of the loopback interface to **0** and its IP address to **10.10.10.10/32**.

   ```
   huawei(config)#interface loopback 0
   huawei(config-if-loopback0)#ip address 10.10.10.10 32
   huawei(config-if-loopback0)#quit
   ```

2. Configure the MPLS LSR-ID. Use the IP address of loopback interface 0 as the LSR ID.

   huawei(config)#**mpls lsr-id 10.10.10.10**

3. Enable MPLS globally.

   Trigger LDP by the IP address of the host to set up an LSP.

   ```
   huawei(config)#mpls
   huawei(config-mpls)#lsp-trigger host
   huawei(config-mpls)#quit
   ```

4. Enable the L2VPN function.

   ```
   huawei(config)#mpls l2vpn
   ```

5. Enable the LDP function globally and enable the split horizon policy.

   ```
   huawei(config)#mpls ldp
   huawei(config-mpls-ldp)#outbound peer all split-horizon
   huawei(config-mpls-ldp)#quit
   ```

Step 2 Configure VLAN, and enable MPLS for VLAN and VLAN interfaces.

1. Add VLAN 4001 for forwarding MPLS packets and add two upstream ports to it.

   ```
   huawei(config)#vlan 4001 smart
   huawei(config)#port vlan 4001 0/9/0
   huawei(config)#port vlan 4001 0/9/1
   ```

2. Set the IP address of VLAN interface 4001 to **10.50.50.50/24** and enable MPLS LDP for the VLAN interface.

   ```
   huawei(config)#interface vlanif 4001
   huawei(config-if-vlanif4001)#ip address 10.50.50.50 24
   huawei(config-if-vlanif4001)#mpls
   huawei(config-if-vlanif4001)#mpls ldp
   huawei(config-if-vlanif4001)#quit
   ```

 Copyright © Huawei Technologies Co., Ltd.

Step 3 Configure routes.

VPLS has no special requirements on routing policy. You can use static route, RIP, or OSPF policy. In the following example, OSPF is used.

Set the OSPF process ID to **100** and OSPF area ID to **1**. In addition, configure the interfaces (VLAN interface and loopback interface) that run OSPF and configure the areas of the interfaces.

```
huawei(config)#ospf 1
huawei(config-ospf-1)#area 100
huawei(config-ospf-1-area-0.0.0.100)#network 10.50.50.0 0.0.0.255
huawei(config-ospf-1-area-0.0.0.100)#network 10.10.10.10 0.0.0.0
huawei(config-ospf-1-area-0.0.0.100)#return
```

Step 4 Configure the remote LDP session.

Configure the remote LDP session from the MA5800 to SPE1 (LSR ID: 10.1.1.1) and SPE2 (LSR ID: 10.2.2.2) respectively. Name the sessions **to_spe1** and **to_spe2** respectively.

```
huawei(config)#mpls ldp remote-peer to_spe1
huawei(config-mpls-ldp-remote-to_spe1)#remote-ip 10.1.1.1
huawei(config-mpls-ldp-remote-to_spe1)#remote-ip auto-dod-request
huawei(config-mpls-ldp-remote-to_spe1)#quit
huawei(config)#mpls ldp remote-peer to_spe2
huawei(config-mpls-ldp-remote-to_spe2)#remote-ip 10.2.2.2
huawei(config-mpls-ldp-remote-to_spe2)#remote-ip auto-dod-request
huawei(config-mpls-ldp-remote-to_spe2)#quit
```

Step 5 Configure a VSI.

1. Add a VSI.

   Create a VSI named **multicast**. Set the signaling protocol to **LDP** and VSI ID to **1**.

   ```
   huawei(config)#vsi multicast
   huawei(config-vsi-multicast)#pwsignal ldp
   huawei(config-vsi-multicast-ldp)#vsi-id 1
   ```

2. (Optional) Disable the suppression of the unknown multicast of the VSI.

   If the VSI unknown multicast suppression is enabled, you need to configure this step. Otherwise, packet loss will occur in the multicast services.

   ```
   huawei(config-vsi-multicast)#undo traffic-suppress multicast
   ```

3. (Optional) Configure the attributes of VSI.

   Configure basic attributes of VSI as required, including the encapsulation type, control word, and MTU value. In the following example, the control word is enabled and default values are used for other parameters.

   ```
   huawei(config-vsi-multicast)#control-word
   ```

Step 6 Configure PWs.

Create two PWs with IDs 1 and 2. Set the service type to **vpls**, the remote IP addresses to **10.1.1.1** and **10.2.2.2** respectively, and encapsulation type to **ethernet tagged**. Enable the control word, and set the receive labels of dynamic PW to **10240** and **10250** respectively.

```
huawei(config)#pw-para pwindex 1
huawei(config-pw-para-index-1)#service-type vpls
huawei(config-pw-para-index-1)#pwid 1
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-pw-para-index-1)#peer-address 10.1.1.1
huawei(config-pw-para-index-1)#pw-type ethernet tagged
huawei(config-pw-para-index-1)#control-word enable
huawei(config-pw-para-index-1)#dyn-receive-label 10240
huawei(config-pw-para-index-1)#quit
huawei(config)#pw-para pwindex 2
huawei(config-pw-para-index-2)#service-type vpls
huawei(config-pw-para-index-2)#pwid 2
huawei(config-pw-para-index-2)#peer-address 10.2.2.2
huawei(config-pw-para-index-2)#pw-type ethernet tagged
huawei(config-pw-para-index-2)#control-word enable
huawei(config-pw-para-index-2)#dyn-receive-label 10250
huawei(config-pw-para-index-2)#quit
```

Step 7 Bind PW and VSI.

Dynamically bind PW1 and PW2 to the VSI named multicast to establish the VPLS PW service.

```
huawei(config)#vsi multicast
huawei(config-vsi-multicast)#pwsignal ldp
huawei(config-vsi-multicast-ldp)#vsi-pw-binding pwindex 1
huawei(config-vsi-multicast-ldp)#vsi-pw-binding pwindex 2
```

Step 8 Bind AC and VSI.

Bind multicast VLAN 100 to the VSI named multicast, so that the multicast service packets of VLAN 100 can be forwarded in VSI.

- The same multicast VLAN cannot be bound to the VPLS instance and PW at the same time.
- The same restrictions are set for binding multicast VLANs and VSI as those for binding unicast VLANs and VSI. That is, when the VPLS encapsulation type is raw, one VIS can only be bound to a multicast VLAN; when the VPLS encapsulation type is tag, one VSI can be bound to multiple multicast VLANs.

```
huawei(config-vsi-multicast)#vsi-ac-binding vlan 100
```

----End

Result

1. A user orders a multicast program. An IGMP packet is transmitted upstream.
2. The MA5800 broadcasts the IGMP packet over two upstream PWs which serve as the multicast upstream ports of the multicast VLAN.
3. On the AGS devices, VPLS exchange is performed on the IGMP packet. The AGS devices learn the VPLS multicast forwarding table and duplicate multicast traffic based on VPLS PW connections.
4. Finally, multicast traffic streams are transmitted over one PW. The MA5800 forwards the multicast traffic streams to the corresponding user port based on the local multicast forwarding table. Then the user can watch the multicast program normally.

9.3.6.3 Configuration Example of the VPLS Enterprise Private Line Service

This topic describes how to configure the enterprise private line service when the VPLS networking is used at the access and aggregation layers.

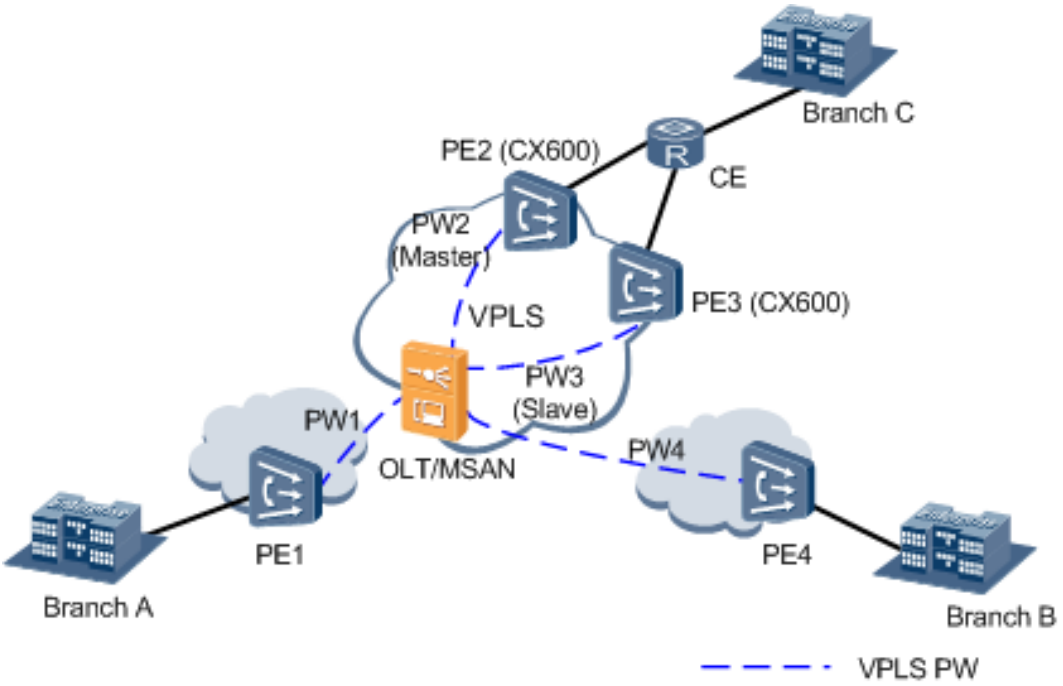
 Copyright © Huawei Technologies Co., Ltd.

Application Context

As shown in **Figure 9-55**, branch offices of an enterprise access the VPLS network through a CE or PE. By deploying VPLS PWs between PEs, the service provider can provide Ethernet-based multipoint services to enterprise users over the MPLS backbone network and achieve emulation of the local area network (LAN). For important branch offices (for example, branch C in the following figure), PW redundancy is configured to provide protection.

In the following figure, as a key node, the OLT/MSAN (MA5800) exchanges data with PEs (PE1-PE4) through VPLS PWs. For important branch offices, a PW protection group is configured to provide protection. PEs are connected to the OLT/MSAN using Spoke PW.

Figure 9-55 Networking for the VPLS enterprise private line service



Prerequisite

QinQ traffic streams have been configured on PE1-PE4 for the private line service of different branch offices.

NOTE

To configure the Ethernet-based enterprise private line service, you must configure QinQ VLAN-based Ethernet traffic streams on PE1-PE4, and perform corresponding configurations on routers for branch offices of the enterprise. The configurations are the same as those for common QinQ VLAN private line service, which are not described here.

Data Plan

Table 9-11 provides the key data plan for the OLT/MSAN (MA5800).

 Copyright © Huawei Technologies Co., Ltd.

Table 9-11 Key data plan

Configuration Item	Data	Remarks	Requirement on PE1-PE4
MPLS	• LSR ID: 10.10.10.10 • VLAN: 4001	MPLS must be enabled at three layers. • MPLS must be enabled globally. • MPLS must be enabled for VLAN. • MPLS must be enabled at VLAN interfaces.	The LSR ID must be unique on the entire network and MPLS must be enabled.
LDP	MPLS LDP is enabled.	MPLS LDP must be enabled at three layers. • MPLS LDP must be enabled globally. • MPLS LDP must be enabled at VLAN interfaces.	MPLS LDP is enabled. The remote LDP session to the MA5800 is configured on PE1-PE4.
Routing protocol	The Open Shortest Path First (OSPF) protocol is used.	Ensure that the Layer 3 interfaces on the MA5800 and those on PE1-PE4 can ping each other.	Layer 3 interfaces and routes are configured on PE1-PE4. Ensure that the Layer 3 interfaces and LSR IDs on PE1-PE4 and those on the MA5800 can ping each other.

 Copyright © Huawei Technologies Co., Ltd.

Configuration Item	Data	Remarks	Requirement on PE1-PE4
VPLS PW	• PW ID: 1, 2, 3, 4 • Service type: vpls • Encapsulation type: ethernet tagged • The control word is enabled. PW protection group: • PW ID: 2, 3 • Working mode: master-slave • The dual receiving function is enabled for PWs.	On the OLT/MSAN, PW1 and PW4 are created for PE1 and PE4 respectively, and PW2 and PW3 are created for PE2 and PE3 respectively. PW2 and PW3 back up each other.	The LDP VPLS is supported, and VPLS PWs to the MA5800 are configured on PE1-PE4, and attributes of PWs are consistent with those on the MA5800.
VSI	• PW1-PW4 are bound to VSI. • Binding mode of PW 1 and PW 4: spoke • VLAN 100 of the private line service is bound to VSI.	VSI binds VLAN and PW to map VLAN to the VPLS domain, so that packets for the enterprise private line service can be broadcast in the VPLS domain.	VSI is configured on PE1-PE4 and the VSI ID must bind the corresponding PW.

Procedure

Step 1 Configure the basic MPLS.

1. Configure a loopback interface.

   Set the ID of the loopback interface to **0** and its IP address to **10.10.10.10/32**.

   ```
   huawei(config)#interface loopback 0
   huawei(config-if-loopback0)#ip address 10.10.10.10 32
   huawei(config-if-loopback0)#quit
   ```

2. Configure the MPLS LSR-ID. Use the IP address of loopback interface 0 as the LSR ID.

   ```
   huawei(config)#mpls lsr-id 10.10.10.10
   ```

 Copyright © Huawei Technologies Co., Ltd.

3. Enable MPLS globally.

   Trigger LDP by the IP address of the host to set up an LSP.

   ```
   huawei(config)#mpls
   huawei(config-mpls)#lsp-trigger host
   huawei(config-mpls)#quit
   ```

4. Enable the L2VPN function.

   ```
   huawei(config)#mpls l2vpn
   ```

5. Enable LDP globally.

   ```
   huawei(config)#mpls ldp
   huawei(config-mpls-ldp)#quit
   ```

Step 2 Configure VLAN, and enable MPLS for VLAN and VLAN interfaces.

1. Add VLAN 4001 for forwarding MPLS packets and add four upstream ports to it.

   ```
   huawei(config)#vlan 4001 smart
   huawei(config)#port vlan 4001 0/9 0
   huawei(config)#port vlan 4001 0/9 1
   huawei(config)#port vlan 4001 0/9 2
   huawei(config)#port vlan 4001 0/9 3
   ```

2. Set the IP address of VLAN interface 4001 to **10.50.50.50/24** and enable MPLS LDP for the VLAN interface.

   ```
   huawei(config)#interface vlanif 4001
   huawei(config-if-vlanif4001)#ip address 10.50.50.50 24
   huawei(config-if-vlanif4001)#mpls
   huawei(config-if-vlanif4001)#mpls ldp
   huawei(config-if-vlanif4001)#quit
   ```

Step 3 Configure routes.

VPLS has no special requirements on routing policy. You can use static route, RIP, or OSPF policy. In the following example, OSPF is used.

Set the OSPF process ID to **100** and OSPF area ID to **1**. In addition, configure the interfaces (VLAN interface and loopback interface) that run OSPF and configure the areas of the interfaces.

```
huawei(config)#ospf 1
huawei(config-ospf-1)#area 100
huawei(config-ospf-1-area-0.0.0.100)#network 10.50.50.0 0.0.0.255
huawei(config-ospf-1-area-0.0.0.100)#network 10.10.10.10 0.0.0.0
huawei(config-ospf-1-area-0.0.0.100)#return
```

Step 4 Configure the remote LDP session.

Configure the remote LDP session from the MA5800 to PE1 (LSR ID: 10.5.5.5), PE2 (LSR ID: 10.2.2.2), PE3 (LSR ID: 10.3.3.3), and PE4 (LSR ID: 10.4.4.4) respectively. Name the sessions **to_pe1**, **to_pe2**, **to_pe3**, and **to_pe4** respectively.

```
huawei(config)#mpls ldp remote-peer to_pe1
huawei(config-mpls-ldp-remote-to_pe1)#remote-ip 10.5.5.5
huawei(config-mpls-ldp-remote-to_pe1)#remote-ip auto-dod-request
huawei(config-mpls-ldp-remote-to_pe1)#quit
huawei(config)#mpls ldp remote-peer to_pe2
huawei(config-mpls-ldp-remote-to_pe2)#remote-ip 10.2.2.2
huawei(config-mpls-ldp-remote-to_pe2)#remote-ip auto-dod-request
huawei(config-mpls-ldp-remote-to_pe2)#quit
huawei(config)#mpls ldp remote-peer to_pe3
huawei(config-mpls-ldp-remote-to_pe3)#remote-ip 10.3.3.3
huawei(config-mpls-ldp-remote-to_pe3)#remote-ip auto-dod-request
huawei(config-mpls-ldp-remote-to_pe3)#quit
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#mpls ldp remote-peer to_pe4
huawei(config-mpls-ldp-remote-to_pe4)#remote-ip 10.4.4.4
huawei(config-mpls-ldp-remote-to_pe4)#remote-ip auto-dod-request
huawei(config-mpls-ldp-remote-to_pe4)#quit
```

Step 5 Configure VSI.

1. Add a VSI.

   Create a VSI named **enterprise_vpn**. Set the signaling protocol to **LDP** and VSI ID to **1**.

   ```
   huawei(config)#vsi enterprise_vpn
   huawei(config-vsi-enterprise_vpn)#pwsignal ldp
   huawei(config-vsi-enterprise_vpn-ldp)#vsi-id 1
   ```

2. (Optional) Configure the attributes of VSI.

   Configure basic attributes of VSI as required, including the encapsulation type, control word, MTU value, and traffic suppression policy. In the following example, the control word is enabled and default values are used for other parameters.

   ```
   huawei(config-vsi-enterprise_vpn)#control-word
   ```

Step 6 Configure PWs.

Create four PWs with IDs 1-4. Set the service type to **vpls**, the remote IP addresses to the IP addresses of PE1-PE4, and encapsulation type to **ethernet tagged**. Enable the control word, and set the receive labels of dynamic PW to **10240**, **10250**, **10260**, and **10270** respectively.

```
huawei(config)#pw-para pwindex 1
huawei(config-pw-para-index-1)#service-type vpls
huawei(config-pw-para-index-1)#pwid 1
huawei(config-pw-para-index-1)#peer-address 10.5.5.5
huawei(config-pw-para-index-1)#pw-type ethernet tagged
huawei(config-pw-para-index-1)#control-word enable
huawei(config-pw-para-index-1)#dyn-receive-label 10240
huawei(config-pw-para-index-1)#quit
huawei(config)#pw-para pwindex 2
huawei(config-pw-para-index-2)#service-type vpls
huawei(config-pw-para-index-2)#pwid 2
huawei(config-pw-para-index-2)#peer-address 10.2.2.2
huawei(config-pw-para-index-2)#pw-type ethernet tagged
huawei(config-pw-para-index-2)#control-word enable
huawei(config-pw-para-index-2)#dyn-receive-label 10250
huawei(config-pw-para-index-2)#quit
huawei(config)#pw-para pwindex 3
huawei(config-pw-para-index-3)#service-type vpls
huawei(config-pw-para-index-3)#pwid 3
huawei(config-pw-para-index-3)#peer-address 10.3.3.3
huawei(config-pw-para-index-3)#pw-type ethernet tagged
huawei(config-pw-para-index-3)#control-word enable
huawei(config-pw-para-index-3)#dyn-receive-label 10260
huawei(config-pw-para-index-3)#quit
huawei(config)#pw-para pwindex 4
huawei(config-pw-para-index-4)#service-type vpls
huawei(config-pw-para-index-4)#pwid 4
huawei(config-pw-para-index-4)#peer-address 10.4.4.4
huawei(config-pw-para-index-4)#pw-type ethernet tagged
huawei(config-pw-para-index-4)#control-word enable
huawei(config-pw-para-index-4)#dyn-receive-label 10270
huawei(config-pw-para-index-4)#quit
```

Step 7 Bind PW and VSI.

Dynamically bind PW 1 and PW 4 in spoke mode, and dynamically bind PW 2 and PW 3 to the VSI named **enterprise_vpn** to establish the VPLS PW service. The

 Copyright © Huawei Technologies Co., Ltd.

spoke mode is used to identify the peer is a user-side PE, and split horizon is not performed between PWs.

```
huawei(config)#vsi enterprise_vpn
huawei(config-vsi-enterprise_vpn)#pwsignal ldp
huawei(config-vsi-enterprise_vpn-ldp)#vsi-pw-binding pwindex 1 spoke
huawei(config-vsi-enterprise_vpn-ldp)#vsi-pw-binding pwindex 2
huawei(config-vsi-enterprise_vpn-ldp)#vsi-pw-binding pwindex 3
huawei(config-vsi-enterprise_vpn-ldp)#vsi-pw-binding pwindex 4 spoke
```

Step 8 Configure VPLS PW protection.

Configure PW2 and PW3 as a PW protection group named **pg_pw**. Set the working mode to master-slave and enable the dual receiving function for PWs. In the protection group enabled with dual receiving, two PWs are always allowed to receive traffic. In this way, when the remote device performs traffic switching, the traffic will not be dropped.

```
huawei(config)#vsi enterprise_vpn
huawei(config-vsi-enterprise_vpn)#pwsignal ldp
huawei(config-vsi-enterprise_vpn-ldp)#protect-group pg_pw
huawei(config-vsi-enterprise_vpn-ldp-group-pg_pw)#pw-protect primary-pw pwindex 2 secondary-pw pwindex 3
huawei(config-vsi-enterprise_vpn-ldp-group-pg_pw)#protect-mode master
huawei(config-vsi-enterprise_vpn-ldp-group-pg_pw)#stream-dual-receiving
```

Step 9 Bind AC and VSI.

Bind QinQ VLAN 100 to the VSI named enterprise_vpn, so that the enterprise private line service packets can be forwarded in the VSI.

```
huawei(config-vsi-enterprise_vpn)#vsi-ac-binding vlan 100
```

----End

Result

As shown in **Figure 9-55**, the private networks distributed in different branches can establish point-to-multipoint communication with each other and various services can be provisioned between these private networks. When a PE which branch C is connected to is faulty, services can be automatically switched to another PE and therefore services are not affected.

9.4 VRF

Virtual route forwarding instance (VRF) is a mechanism in which a device works as multiple virtual routing devices. After the Layer 3 interfaces of the device are divided into different VRFs, multiple route forwarding instances can be emulated on the device.

9.4.1 Introduction to VRF

Definition

The key of the virtual route forwarding (VRF) lies in VR, which is equivalent to implementing the functions of multiple virtual routing devices one routing device. VRF is suitable for GPON users.

 Copyright © Huawei Technologies Co., Ltd.

Purpose

The MA5800 categorizes VRFs by VLANs to provide L3VPN solutions. All the packets or related protocols on the Layer 3 interface of a VRF are processed only in this VRF, which is unrelated to other VRFs. In this way, the services or users can be isolated, and the IP addresses can be saved.

Multiple virtual routing devices can be created on the MA5800. That is, multiple L3VPNs can be established to implement the Layer 3 isolation and independent packet forwarding among different VRFs. Moreover, in different VRFs, the IP address can be reused, and also DHCP relay multi-instances, routing multi-instances, and independent route forwarding tables are supported.

VRF has two application scenarios:

- When the triple play service is provisioned to GPON access users, different services are isolated from each other by VRF, and all services of the device are carried and go upstream by the same physical link. One VR can be bound to each VLANIF interface, and the upstream port belongs to multiple Layer 3 interfaces. Different VRs are bound to different VLANIF interfaces, and each VR forwards data according to the route learned by this VR.
- When the triple play service is provisioned to GPON access users, different services are isolated from each other by VRF, and all services of the device are carried and go upstream by two or more physical links. The links in this case are in the Layer 3 mode, and different services are isolated from each other by VRF.

The difference of the two scenarios is that dual or multiple links are adopted for upstream transmission in scenario 2, where the effect of different VRs going different "ways" is more vivid.

9.4.2 VRF Principle

VRF Compatibility

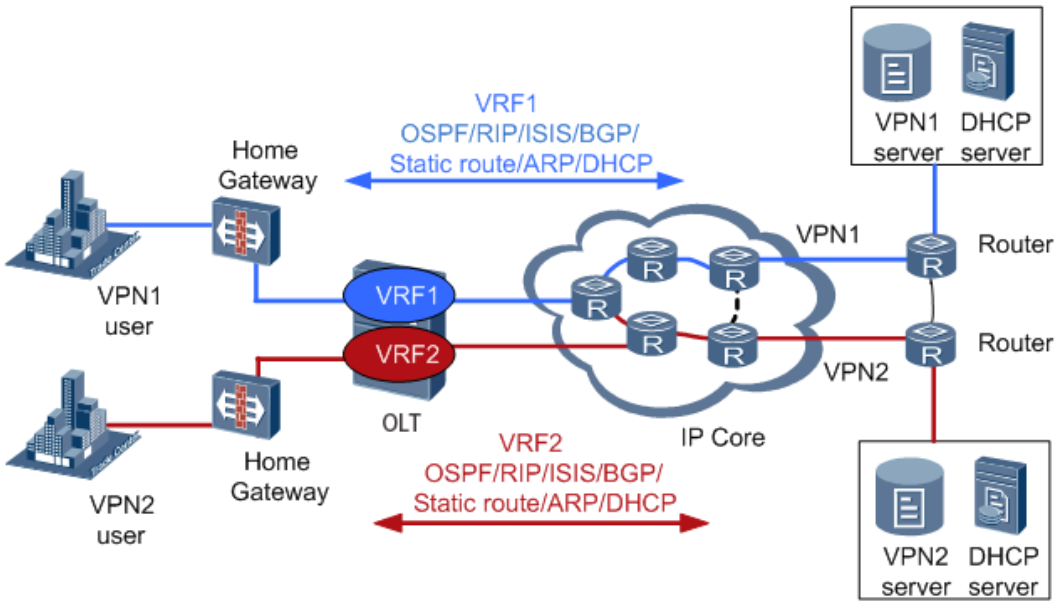
The VRF architecture is compatible with the virtual private routed network (VPRN) architecture as defined in RFC2764.

VRF Architecture

VRF is an architecture of IP networks, as shown in **Figure 9-56**. When users are isolated by service types or ISPs, or the users of different VPNs are prohibited from communicating with each other, multiple L3VPNs must be established in an IP network.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-56 VRF architecture



The MA5800 VRF supports the following functions:

- Creating a VRF instance

  You can create a VRF instance and set the name of the VRF as the reference flag through the CLI.
- Adding the VLAN Layer 3 interface and the loopback interface into a VRF instance
  - The MA5800 differentiates VRFs by the VLAN Layer 3 interfaces. A VRF contains one or more VLAN Layer 3 interfaces. When receiving or transmitting packets, any VLAN Layer 3 interface that belongs to the VRF must use the Layer 3 route forwarding table of the VRF. Moreover, the packets in the VRF must be forwarded between these VLAN Layer 3 interfaces and cannot be forwarded to any other VLAN Layer 3 interfaces that do not belong to the VRF.
  - After a loopback interface is bound with the VRF instance, the loopback interface can process all the routing protocols in the VRF.
  - The IP addresses configured in the VLAN Layer 3 interfaces of different VRFs can be identical, but the IP addresses in the same VRF cannot be identical.
- Isolating ARP in a VRF

  The ARP in different VRFs is isolated, but the user IP addresses in different VRFs can be identical.
- Supporting independent ISIS, OSPF, RIP, or BGP routing protocol process for different VRFs
- Supporting the Layer 3 DHCP relay or DHCP proxy in a VRF

  The MA5800 supports the DHCP configuration based on the VLAN to implement the DHCP relay or DHCP proxy function in the VRF.
- Supporting the ping and trace route functions in a VRF
  - Ping and trace route are the basic network maintenance means.

 Copyright © Huawei Technologies Co., Ltd.

- The ping function is used to check the connectivity and reachability of a remote host by sending the ping packets to the host.
- The trace route function is used to check the network connectivity and locate the network faults by testing the route that the data packets pass through from the host to the destination.

9.4.3 Configuring IPv4 in VPN

This topic describes how to categorize virtual private network (VPN) instances by VLANs, and realize the virtual IPv4 static route forwarding in different VPN instances.

Context

A VPN instance is also called a VPN Routing and Forwarding (VRF) table. VRF is a Layer 3 virtual private network (L3 VPN). VRF is a mechanism in which a device works as multiple virtual routing devices. After the Layer 3 interfaces of the device are divided into different VRFs, multiple route forwarding instances can be emulated on the device.

Multiple virtual routing devices can be created on the access node. That is, multiple L3VPNs can be established to implement the Layer 3 isolation and independent packet forwarding among different VRFs. The access node supports the following VRF functions:

- In different VRF instances, the IP address can be reused. It means that the IP addresses of the Layer 3 interfaces which belong to different VRF instances can be the same.
- The ping and trace route functions are supported in a VRF.
- The users of different VRF instances can obtain the IP addresses through the Dynamic Host Control Protocol (DHCP) relay or the DHCP proxy.
- The static routes and the dynamic routes in a VRF instance do not affect each other, and the routing entry in each VRF instance supports the routing function independently.

Networking

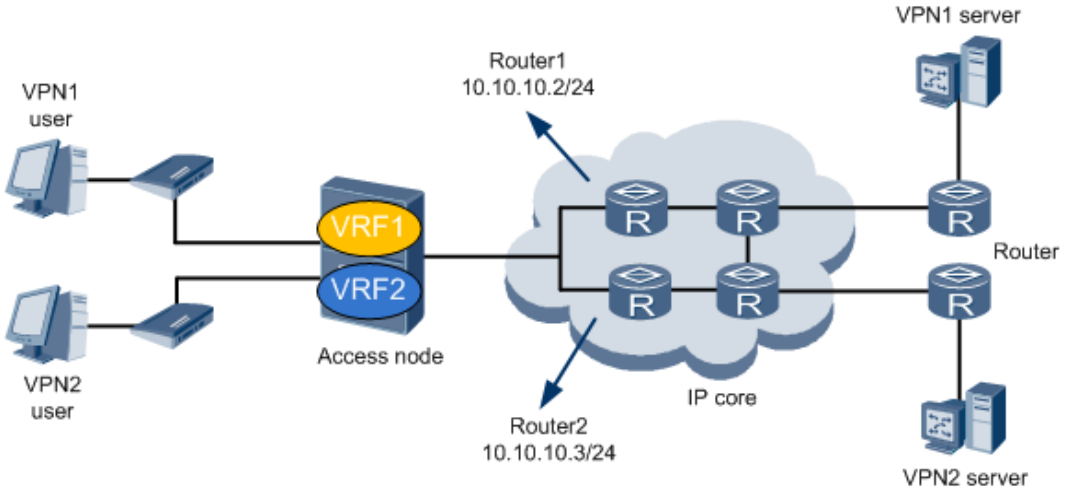
Figure 9-57 shows an example network for configuring IPv4 in VPN.

The access node categorizes VRF instances by VLANs to provide L3 VPN solutions. In this example, VPN instance VRF1 is categorized by virtual local area network (VLAN) 200, and IPv4 static routes are added in the virtual route forwarding entries of VRF1. The access node selects the routes for the users of VPN1 by querying the routing entries of VRF1. Similarly, VPN instance VRF2 is categorized by VLAN 300 and is used to select the routes for the users of VPN2. The access node implements the Layer 3 isolation and independent packet forwarding through different VRF instances.

This example describes how to configure the function of virtual static route forwarding by adding IPv4 static routes application on the instance. The function of virtual dynamic route forwarding can be realized by enabling the process of the dynamic routing protocols such as the open shortest path first (OSPF), Routing Information Protocol (RIP), intermediate system to intermediate system (IS-IS), and Border Gateway Protocol (BGP) in a VRF instance.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-57 Example network for configuring IPv4 in VPN



Data Plan

Table 9-12 provides the data plan for configuring IPv4 in VPN.

Table 9-12 Data plan for configuring IPv4 in VPN

Item	Data
VRF1 (for VPN1)	Name of the VPN instance: vpn1 Route distinguisher (RD) of the VPN instance: 100:1
	Upstream port: 0/9/0 VLAN ID: 200 VLAN type: Smart VPN1 user: • Gigabit-capable passive optical network (GPON) port: 0/2/0 • ONT ID: 0 • GEM Port ID: 0
	IP address of the Layer 3 interface of VLAN 200: 10.10.10.1/24 IP address of router1: 10.10.10.2/24 IP address of the VPN1 server: 10.10.20.1/24
VRF2 (for VPN2)	Name of the VPN instance: vpn2 RD of the VPN instance: 100:2

 Copyright © Huawei Technologies Co., Ltd.

Item	Data
	Upstream port: 0/9/0 VLAN ID: 300 VLAN type: Smart VPN2 user: • GPON port: 0/2/1 • ONT ID: 1 • GEM Port ID: 1
	IP address of the Layer 3 interface of VLAN 300: 10.10.10.1/24 IP address of router2: 10.10.10.3/24 IP address of the VPN2 server: 10.10.30.1/24

Procedure

- Configure VRF1 (for VPN1).
  a. Create a VPN instance of the IPv4 address family.

```
huawei(config)#ip vpn-instance vpn1
huawei(config-vpn-instance-vpn1)#ipv4-family
```

  b. Configure the RD of the VPN instance.

```
huawei(config-vpn-instance-vpn1-af-ipv4)#route-distinguisher 100:1
huawei(config-vpn-instance-vpn1-af-ipv4)#quit
huawei(config-vpn-instance-vpn1)#quit
```

  c. Create a smart VLAN and add the upstream port and the service port to it.

```
huawei(config)#vlan 200 smart
huawei(config)#port vlan 200 0/9 0
huawei(config)#service-port vlan 200 gpon 0/2/0 ont 0 gemport 0 multi-service user-8021p 0 user-vlan 200 rx-cttr 5 tx-cttr 5
```

  d. Associate the Layer 3 interface with the VPN instance.

```
huawei(config)#interface vlanif 200
huawei(config-if-vlanif200)#ip binding vpn-instance vpn1
  Info: All IPv4 and IPv6 related configurations on this interface are removed.
```

  e. Configure the IP address of the VLAN Layer 3 interface.

```
huawei(config-if-vlanif200)#ip address 10.10.10.1 24
huawei(config-if-vlanif200)#quit
```

  f. Configure the IPv4 static route.

```
huawei(config)#ip route-static vpn-instance vpn1 10.10.20.0 24 10.10.10.2
```

  g. Save the data.

```
huawei(config)#save
```

- Configure VRF2 (for VPN2).
  a. Create a VPN instance of the IPv4 address family.

```
huawei(config)#ip vpn-instance vpn2
huawei(config-vpn-instance-vpn2)#ipv4-family
```

  b. Configure the RD of the VPN instance.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-vpn-instance-vpn2-af-ipv4)#route-distinguisher 100:2
huawei(config-vpn-instance-vpn2-af-ipv4)#quit
huawei(config-vpn-instance-vpn2)#quit
```

c. Create a smart VLAN and add the upstream port and the service port to it.

```
huawei(config)#vlan 300 smart
huawei(config)#port vlan 300 0/9 0
huawei(config)#service-port vlan 300 gpon 0/2/1 ont 1 gemport 1 multi-service user-8021p 0 user-vlan 300 rx-cttr 6 tx-cttr 6
```

d. Associate the Layer 3 interface with the VPN instance.

```
huawei(config)#interface vlanif 300
huawei(config-if-vlanif300)#ip binding vpn-instance vpn2
  Info: All IPv4 and IPv6 related configurations on this interface are removed.
```

e. Configure the IP address of the VLAN Layer 3 interface.

```
huawei(config-if-vlanif300)#ip address 10.10.10.1 24
huawei(config-if-vlanif300)#quit
```

f. Configure the IPv4 static route.

```
huawei(config)#ip route-static vpn-instance vpn2 10.10.30.0 24 10.10.10.3
```

g. Save the data.

```
huawei(config)#save
```

----End

Result

Run the **display ip vpn-instance** command to query the VPN configuration.

```
huawei(config)#display ip vpn-instance
{ <cr>|import-vt<K>|interface<K>|tunnel-info<K>|verbose<K>|vpn-instance-name<S><
Length 1-31>|vpn1|vpn2||<K> }:

  Command:
        display ip vpn-instance
 Total VPN-Instances configured      : 2
 Total IPv4 VPN-Instances configured : 2
 Total IPv6 VPN-Instances configured : 0

  VPN-Instance Name            RD                Address-family
  vpn1                         100:1             IPv4
  vpn2                         100:2             IPv4
```

Run the following commands to verify that the VRF instances are configured successfully. The two IPv4 static routes are added to the IP routing table of VPN1 and VPN2.

```
huawei(config)#display ip routing-table vpn-instance vpn1
{ <cr>|ip-prefix<K>|prefix<I><X.X.X.X>|protocol<K>|route-number<K>|statistics<K>|table-name<K>|time-range<K>|verbose<K>||<K> }:

  Command:
        display ip routing-table vpn-instance vpn1
Route Flags: R - relay, D - download to fib, T - to vpn-instance, B - black hole route
------------------------------------------------------------------------------
Routing Tables: vpn1
         Destinations : 3        Routes : 3

Destination/Mask    Proto  Pre  Cost      Flags NextHop         Interface

     10.10.10.0/24  Direct 0    0           D   10.10.10.1      vlanif200
     10.10.10.1/32  Direct 0    0           D   127.0.0.1       InLoopBack0
     10.10.20.0/24  Static 60   0           D   10.10.10.2      vlanif200
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#display ip routing-table vpn-instance vpn2
{ <cr>|ip-prefix<K>|prefix<I><X.X.X.X>|protocol<K>|route-number<K>|statistics<K>|table-name<K>|time-
range<K>|verbose<K>||<K> }:

  Command:
        display ip routing-table vpn-instance vpn2
Route Flags: R - relay, D - download to fib, T - to vpn-instance, B - black hole
route
------------------------------------------------------------------------------
Routing Tables: vpn2
       Destinations : 3        Routes : 3

Destination/Mask    Proto  Pre  Cost      Flags NextHop         Interface

    10.10.10.0/24  Direct 0    0           D   10.10.10.1      vlanif300
    10.10.10.1/32  Direct 0    0           D   127.0.0.1       InLoopBack0
    10.10.30.0/24  Static 60   0           D   10.10.10.3      vlanif300
```

Run the **ping** and **tracert** commands to check the VPN connectivity.

The MA5800 categorizes VRF instances by VLANs to provide L3 VPN solutions, realizing the Layer 3 isolation of users or services.

- For the users of VPN1, the MA5800 selects the routes by querying the routing entries of VPN1. For example, for the packets to be sent to the VPN1 server (with IP address 10.10.20.1), the MA5800 selects its next hop router (with IP address 10.10.10.2) to forward the packets.
- For the users of VPN2, the MA5800 selects the routes by querying the routing entries of VPN2. For example, for the packets to be sent to the VPN2 server (with IP address 10.10.30.1), the MA5800 selects its next hop router (with IP address 10.10.10.3) to forward the packets.
- For the users outside the VPNs, the route to the VPN1 server or the VPN2 server is not available.

Configuration File

Only the configuration files related to the VPN are listed.

```
ip vpn-instance vpn1
ipv4-family
route-distinguisher 100:1
quit
quit
ip vpn-instance vpn2
ipv4-family
route-distinguisher 100:2
quit
quit
interface vlanif200
ip binding vpn-instance vpn1
ip address 10.10.10.1/24
quit
interface vlanif300
ip binding vpn-instance vpn2
ip address 10.10.10.1/24
quit
ip route-static vpn-instance vpn1 10.10.20.0 24 10.10.10.2
ip route-static vpn-instance vpn2 10.10.30.0 24 10.10.10.3
```

9.4.4 Configuring IPv6 in VPN

This topic describes how to categorize virtual private network (VPN) instances by virtual local area networks (VLANs), and implement the virtual IPv6 static route forwarding in different VPN instances.

 Copyright © Huawei Technologies Co., Ltd.

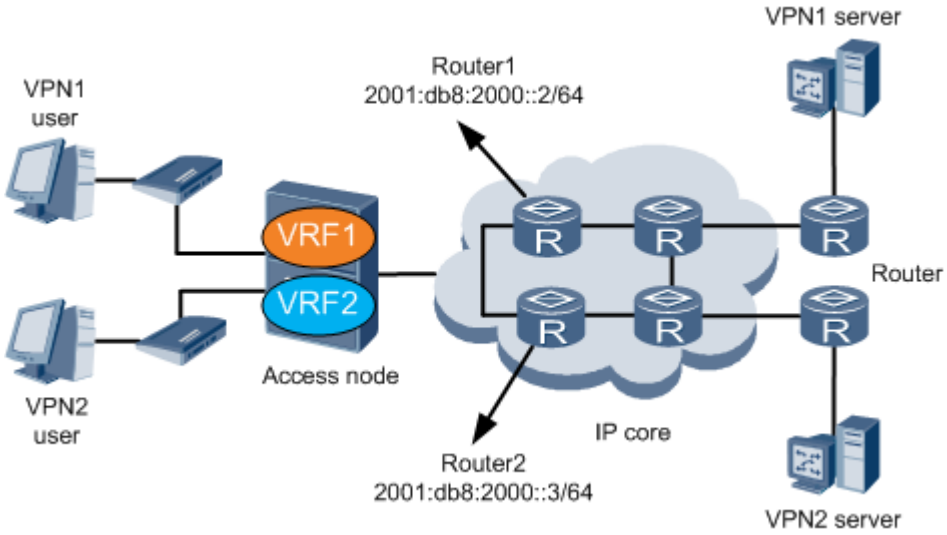
Networking

The access node supports virtual route forwarding (VRF) in the IPv6 network. The VRF principle and functions in the IPv6 network are the same as those in the IPv4 network. **9.4.4 Configuring IPv6 in VPN** shows an example network for configuring IPv6 in VPN.

The access node categorizes VRF instances by VLANs to provide L3 VPN solutions. In this example, VPN instance VRF1 is categorized by virtual local area network (VLAN) 200, and IPv6 static routes are added in the virtual route forwarding entries of VRF1. The access node selects the routes for the users of VPN1 by querying the routing entries of VRF1. Similarly, VPN instance VRF2 is categorized by VLAN 300 and is used to select the routes for the users of VPN2. The access node implements the Layer 3 isolation and independent packet forwarding through different VRF instances.

This example describes how to configure the function of virtual static route forwarding by adding IPv6 static routes to a VRF routing table. In addition, virtual dynamic route forwarding can be implemented in a VPN by using BGP4+.

Figure 9-58 IPv6 in VPN networking diagram



Data Plan

Table 9-13 provides the data plan for configuring IPv6 in VPN.

Table 9-13 Data plan for configuring IPv6 in VPN

Item	Data
VRF1 (for VPN1)	Name of the VPN instance: vpn1 Route distinguisher (RD) of the VPN instance: 100:1

 Copyright © Huawei Technologies Co., Ltd.

Item	Data
	Upstream port: 0/9/0 VLAN ID: 200 VLAN type: Smart VPN1 user: ● Gigabit-capable passive optical network (GPON) port: 0/2/0 ● ONT ID: 0 ● GEM Port ID: 0
	IPv6 address of the Layer 3 interface of VLAN 200: 2001:db8:2000::1/64 IPv6 address of router1: 2001:db8:2000::2/64 IPv6 address of the VPN1 server: 2001:db8:2001::1/64
VRF2 (for VPN2)	Name of the VPN instance: vpn2 Route distinguisher (RD) of the VPN instance: 100:2
	Upstream port: 0/9/0 VLAN ID: 300 VLAN type: Smart VPN2 user: ● GPON port: 0/2/1 ● ONT ID: 1 ● GEM Port ID: 1
	IPv6 address of the Layer 3 interface of VLAN 300: 2001:db8:2000::1/64 IPv6 address of router2: 2001:db8:2000::3/64 IPv6 address of the VPN2 server: 2001:db8:2002::1/64

Procedure

- Configure VRF1 (for VPN1).
  a. Create a VPN instance of the IPv6 address family.
  ```
  huawei(config)#ip vpn-instance vpn1
  huawei(config-vpn-instance-vpn1)#ipv6-family
  ```
  b. Configure the RD of the VPN instance.
  ```
  huawei(config-vpn-instance-vpn1-af-ipv6)#route-distinguisher 100:1
  ```
  c. (Optional) Configure the IPv6 routing specifications of the VPN instance.
  ```
  huawei(config-vpn-instance-vpn1-af-ipv6)#prefix limit 1000 simply-alert
  huawei(config-vpn-instance-vpn1-af-ipv6)#quit
  huawei(config-vpn-instance-vpn1)#quit
  ```
  d. Create a smart VLAN and add the upstream port and the service port to it.
  ```
  huawei(config)#vlan 200 smart
  huawei(config)#port vlan 200 0/9 0
  ```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#service-port vlan 200 gpon 0/2/0 ont 0 gemport 0 multi-service user-8021p
0 user-vlan 200 rx-cttr 5 tx-cttr 5
```

e. Associate the Layer 3 interface with the VPN instance.

```
huawei(config)#interface vlanif 200
huawei(config-if-vlanif200)#ip binding vpn-instance vpn1
  Info: All IPv4 and IPv6 related configurations on this interface are removed.
```

f. Configure the IPv6 address of the VLAN Layer 3 interface.

```
huawei(config-if-vlanif200)#ipv6 enable
huawei(config-if-vlanif200)#ipv6 address 2001:db8:2000::1 64
huawei(config-if-vlanif200)#quit
```

g. Configure the IPv6 static route.

```
huawei(config)#ipv6 route-static vpn-instance vpn1 2001:db8:2001:: 64 2001:db8:2000::2
```

h. Save the data.

```
huawei(config)#save
```

- Configure VRF2 (for VPN2).
  a. Create a VPN instance of the IPv6 address family.

```
huawei(config)#ip vpn-instance vpn2
huawei(config-vpn-instance-vpn2)#ipv6-family
```

  b. Configure the RD of the VPN instance.

```
huawei(config-vpn-instance-vpn2-af-ipv6)#route-distinguisher 100:2
```

  c. (Optional) Configure the IPv6 routing specifications of the VPN instance.

```
huawei(config-vpn-instance-vpn2-af-ipv6)#prefix limit 1000 simply-alert
huawei(config-vpn-instance-vpn2-af-ipv6)#quit
huawei(config-vpn-instance-vpn2)#quit
```

  d. Create a smart VLAN and add the upstream port and the service port to it.

```
huawei(config)#vlan 300 smart
huawei(config)#port vlan 300 0/9 0
huawei(config)#service-port vlan 300 gpon 0/2/1 ont 1 gemport 1 multi-service user-8021p
0 user-vlan 300 rx-cttr 6 tx-cttr 6
```

  e. Associate the Layer 3 interface with the VPN instance.

```
huawei(config)#interface vlanif 300
huawei(config-if-vlanif300)#ip binding vpn-instance vpn2
  Info: All IPv4 and IPv6 related configurations on this interface are removed.
```

  f. Configure the IPv6 address of the VLAN Layer 3 interface.

```
huawei(config-if-vlanif300)#ipv6 enable
huawei(config-if-vlanif300)#ipv6 address 2001:db8:2000::1 64
huawei(config-if-vlanif300)#quit
```

  g. Configure the IPv6 static route.

```
huawei(config)#ipv6 route-static vpn-instance vpn2 2001:db8:2002:: 64 2001:db8:2000::3
```

  h. Save the data.

```
huawei(config)#save
```

----End

Result

Run the **display ip vpn-instance** command to query the VPN configurations.

```
huawei(config)#display ip vpn-instance
{ <cr>|import-vt<K>|interface<K>|tunnel-info<K>|verbose<K>|vpn-instance-name<S><
Length 1-31>|vpn1|vpn2|<K> }:

 Command:
      display ip vpn-instance
```

 Copyright © Huawei Technologies Co., Ltd.

```
Total VPN-Instances configured : 2
Total IPv4 VPN-Instances configured : 0
Total IPv6 VPN-Instances configured : 2

 VPN-Instance Name              RD                  Address-family
 vpn1                        100:1              IPv6
 vpn2                        100:2              IPv6
```

Run the following commands to verify that the configurations are successful and the IPv6 static route is added to the IPv6 routing table of VPN1 and VPN2.

```
huawei(config)#display ipv6 routing-table vpn-instance vpn1
{ <cr>|ipv6-prefix<K>|ipv6prefix<IPv6><X:X::X:X>|protocol<K>|route-number<K>|sta
tistics<K>|time-range<K>|verbose<K>||<K> }:

  Command:
          display ipv6 routing-table vpn-instance vpn1
Routing Table : vpn1
        Destinations : 4        Routes : 4

 Destination  : 2001:db8:2000::                PrefixLength : 64
 NextHop      : 2001:db8:2000::1                Preference   : 0
 Cost         : 0                          Protocol     : Direct
 RelayNextHop : ::                             TunnelID     : 0x0
 Interface    : vlanif200                      Flags        : D

 Destination  : 2001:db8:2000::1                PrefixLength : 128
 NextHop      : ::1                          Preference   : 0
 Cost         : 0                          Protocol     : Direct
 RelayNextHop : ::                             TunnelID     : 0x0
 Interface    : InLoopBack0                      Flags        : D

 Destination  : 2001:db8:2001::                PrefixLength : 64
 NextHop      : 2001:db8:2000::2                Preference   : 60
 Cost         : 0                          Protocol     : Static
 RelayNextHop : ::                             TunnelID     : 0x0
 Interface    : vlanif200                      Flags        : RD

 Destination  : FE80::                         PrefixLength : 10
 NextHop      : ::                           Preference   : 0
 Cost         : 0                          Protocol     : Direct
 RelayNextHop : ::                             TunnelID     : 0x0
 Interface    : null0                        Flags        : D
huawei(config)#display ipv6 routing-table vpn-instance vpn2
{ <cr>|ipv6-prefix<K>|ipv6prefix<IPv6><X:X::X:X>|protocol<K>|route-number<K>|sta
tistics<K>|time-range<K>|verbose<K>||<K> }:

  Command:
          display ipv6 routing-table vpn-instance vpn2
Routing Table : vpn2
        Destinations : 4        Routes : 4

 Destination  : 2001:db8:2000::                PrefixLength : 64
 NextHop      : 2001:db8:2000::1                Preference   : 0
 Cost         : 0                          Protocol     : Direct
 RelayNextHop : ::                             TunnelID     : 0x0
 Interface    : vlanif200                      Flags        : D

 Destination  : 2001:db8:2000::1                PrefixLength : 128
 NextHop      : ::1                          Preference   : 0
 Cost         : 0                          Protocol     : Direct
 RelayNextHop : ::                             TunnelID     : 0x0
 Interface    : InLoopBack0                      Flags        : D

 Destination  : 2001:db8:2002::                PrefixLength : 64
 NextHop      : 2001:db8:2000::3                Preference   : 60
 Cost         : 0                          Protocol     : Static
 RelayNextHop : ::                             TunnelID     : 0x0
 Interface    : vlanif300                      Flags        : RD
```

```
Destination  : FE80::                    PrefixLength : 10
NextHop      : ::                       Preference   : 0
Cost         : 0                        Protocol     : Direct
RelayNextHop : ::                       TunnelID     : 0x0
Interface    : null0                    Flags        : D
```

Run the **ping ipv6** and **tracert ipv6** commands to check the VPN connectivity.

The MA5800 categorizes VRF instances by VLANs to provide L3 VPN solutions, realizing the Layer 3 isolation of users or services.

- For the users of VPN1, the MA5800 selects the routes by querying the routing entries of VPN1. For example, for the packets to be sent to the VPN1 server (with IPv6 address 2001:db8:2001::1), the MA5800 selects its next hop router (with IPv6 address 2001:db8:2000::2) to forward the packets.
- For the users of VPN2, the MA5800 selects the routes by querying the routing entries of VPN2. For example, for the packets to be sent to the VPN2 server (with IPv6 address 2001:db8:2002::1), the MA5800 selects its next hop router (with IPv6 address 2001:db8:2000::3) to forward the packets.
- For the users outside the VPNs, the route to the VPN1 server or the VPN2 server is not available.

Configuration File

Only the configuration files related to the VPN are listed.

```
ip vpn-instance vpn1
ipv6-family
route-distinguisher 100:1
routing-table limit 1000 simply-alert
prefix limit 1000 simply-alert
quit
quit
ip vpn-instance vpn2
ipv6-family
route-distinguisher 100:2
routing-table limit 1000 simply-alert
prefix limit 1000 simply-alert
quit
quit
interface vlanif200
ip binding vpn-instance vpn1
ipv6 enable
ipv6 address 2001:db8:2000::1/64
quit
interface vlanif300
ip binding vpn-instance vpn2
ipv6 enable
ipv6 address 2001:db8:2000::1/64
quit
ipv6 route-static vpn-instance vpn1 2001:db8:2001:: 64 2001:db8:2000::2
ipv6 route-static vpn-instance vpn2 2001:db8:2002:: 64 2001:db8:2000::3
```

9.5 BGP/MPLS IP VPN

 Copyright © Huawei Technologies Co., Ltd.

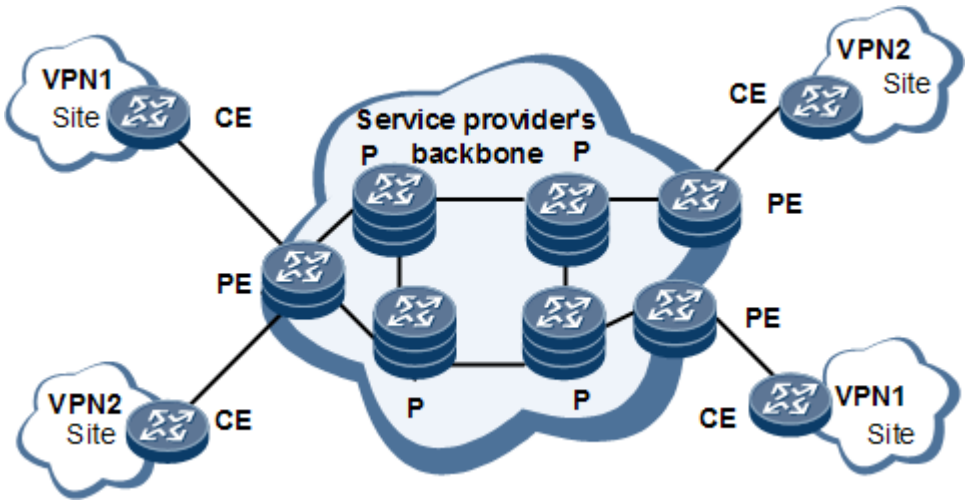
9.5.1 Introduction to BGP/MPLS IP VPN

Definition

A BGP/MPLS IP VPN is a Layer 3 virtual private network (L3VPN), which uses BGP to advertise VPN routes and uses MPLS to forward VPN packets on the IP backbone networks of service providers (SPs).

Figure 9-59 shows the basic model of BGP/MPLS IP VPN.

Figure 9-59 BGP/MPLS IP VPN



As shown in **Figure 9-59**, a BGP/MPLS IP VPN consists of the following roles:

- CE: An edge device on a customer network. A CE provides interfaces that are directly connected to the SP network. A CE can be a router, a switch, or a host. Usually, a CE is unaware of the VPN and does not need to support MPLS.
- PE: An edge device on an SP network. A PE is directly connected to a CE. On an MPLS network, PEs process all VPN services. The requirements on the performance of PEs are rather high.
- P: A backbone device on an SP network. A P is not directly connected to a CE. Ps only need to possess basic MPLS forwarding capabilities and do not maintain VPN information.

PEs and Ps are managed by SPs. CEs are managed by users, except that the users trust SPs with the management rights.

A PE can connect to multiple CEs. A CE can connect to multiple PEs of the same SP or of different SPs.

Purpose

- MPLS seamlessly integrates the flexibility of IP routing and simplicity of ATM label switching. A connection-oriented control plane is introduced into an MPLS IP network, which enriches the means of managing and operating the

 Copyright © Huawei Technologies Co., Ltd.

network. On IP networks, MPLS TE has become an important tool in managing network traffic, reducing network congestion, and ensuring QoS.

The VPNs using MPLS IP networks as the backbone networks are highly valued by carriers, and have become an important means of providing value-added services.

- Unlike the IGP, BGP focuses on controlling route transmission and choosing optimal routes instead of discovering and calculating routes. VPNs use public networks to transmit VPN data, and the public networks use an IGP to discover and calculate their routes. The key to constructing a VPN is to control the transmission of VPN routes and choose the optimal routes between two PEs.

  BGP uses TCP (with port number 179) as the transport layer protocol, enhancing transmission reliability. VPN routes can be directly exchanged between two PEs with routers located between them.

  BGP can append any information to a route as optional BGP attributes. The information is transparently forwarded by BGP devices that cannot identify those attributes. Therefore, VPN routes can be conveniently transmitted between PEs.

  When routes are updated, BGP sends only updated routes rather than all routes. This implementation saves the bandwidth consumed by route transmission, making the transmission of a great number of routes over a public network possible.

  As an Exterior Gateway Protocol (EGP), BGP is best suited for VPNs that cross the networks of multiple carriers.

9.5.2 BGP/MPLS IP VPN Reference Standards and Protocols

The following table lists the references.

Document No.	Document Name	Protocol Compliance
RFC 2764	A Framework for IP Based Virtual Private Networks	Fully compliant
RFC 2917	A Core MPLS IP VPN Architecture	Fully compliant
RFC 3392	Capabilities Advertisement with BGP-4	Fully compliant
RFC 4026	Provider Provisioned Virtual Private Network (VPN) Terminology	Fully compliant
RFC 4271	BGP-4	Fully compliant
RFC 4364	BGP/MPLS IP Virtual Private Networks (VPNs)	Partially compliant. The Carrier's Carrier is not supported.
RFC 4382	MPLS/BGP Layer 3 Virtual Private Network (VPN) Management Information Base	Fully compliant

 Copyright © Huawei Technologies Co., Ltd.

Document No.	Document Name	Protocol Compliance
RFC 4577	OSPF as the Provider/Customer Edge Protocol for BGP/MPLS IP Virtual Private Networks (VPNs)	Fully compliant
RFC 4659	BGP-MPLS VPN Extension for IPv6 VPN	Partially compliant. The Carrier's Carrier is not supported.
RFC 4760	Multiprotocol Extensions for BGP-4	Fully compliant
RFC 5668	4-Octet AS Specific BGP Extended Community	Fully compliant

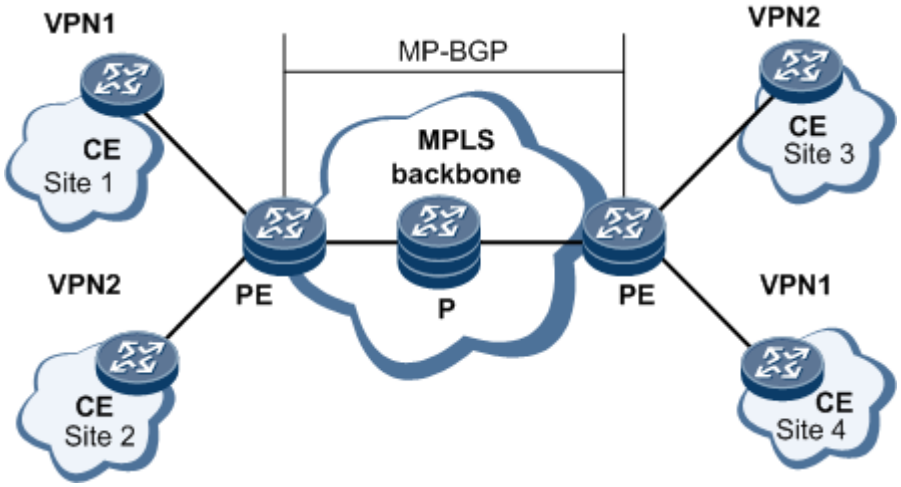
9.5.3 Basic BGP/MPLS IP VPN

Definition

A basic BGP/MPLS IP VPN is an L3VPN network that covers only one carrier's network, which is an MPLS backbone network that does not span multiple ASs, as shown in **Figure 9-60**. A basic BGP/MPLS IP VPN has the following characteristics:

- Transmits packets using extended BGP.
- Encapsulates and transmits VPN packets over MPLS LSPs serving as public network tunnels.
- Allows a device that can play PE, P, and CE roles to play only one role at a time.

Figure 9-60 Basic BGP/MPLS IP VPN networking



 Copyright © Huawei Technologies Co., Ltd.

Benefits

BGP/MPLS IP VPN offers the following benefits:

- Enables users to communicate with each other over networks of geographically different regions.
- Ensures the security of VPN user data during transmission over the public network.

9.5.4 Related Concepts for BGP/MPLS IP VPN

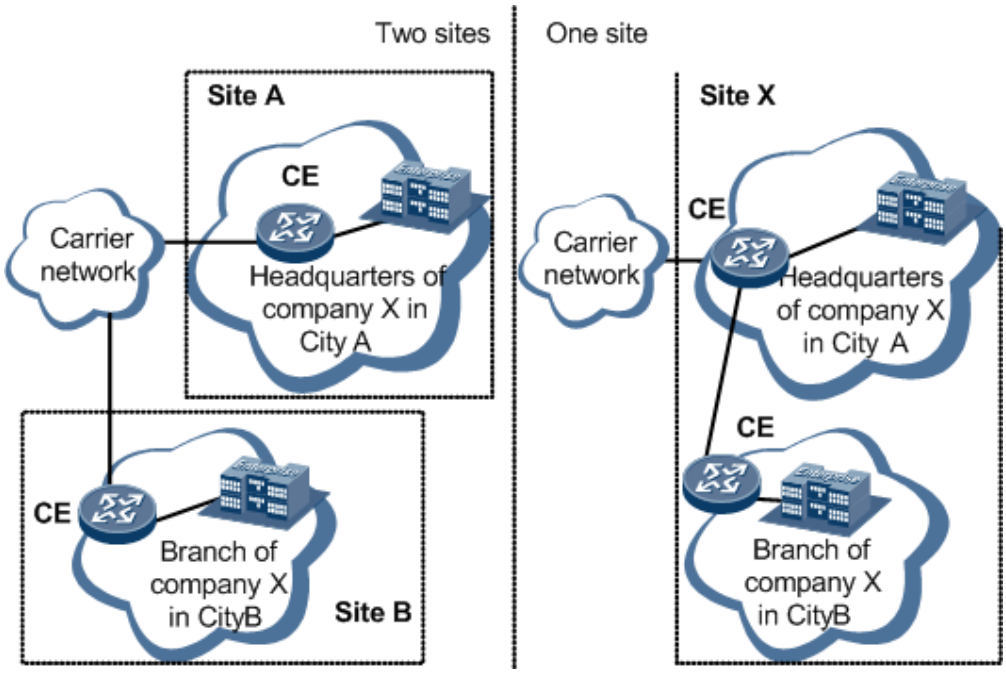
Site

The concept of "site" is frequently mentioned in the VPN technology. The following describes a site from different aspects:

- A site is a group of IP systems that can communicate without using carrier networks.

  As shown in **Figure 9-61**, on the networks of the left side, the headquarters network of company X in City A is a site; the branch network of company X in City B is another site. IP devices within each site can communicate without using the SP network.

  Figure 9-61 Schematic diagram of sites



- Sites are classified based on the topological relationships between devices rather than the geographical locations of devices, although devices in a site are geographically adjacent to each other in general. If two geographically separated IP systems are connected over a leased line instead of a carrier network, the two systems compose a site.

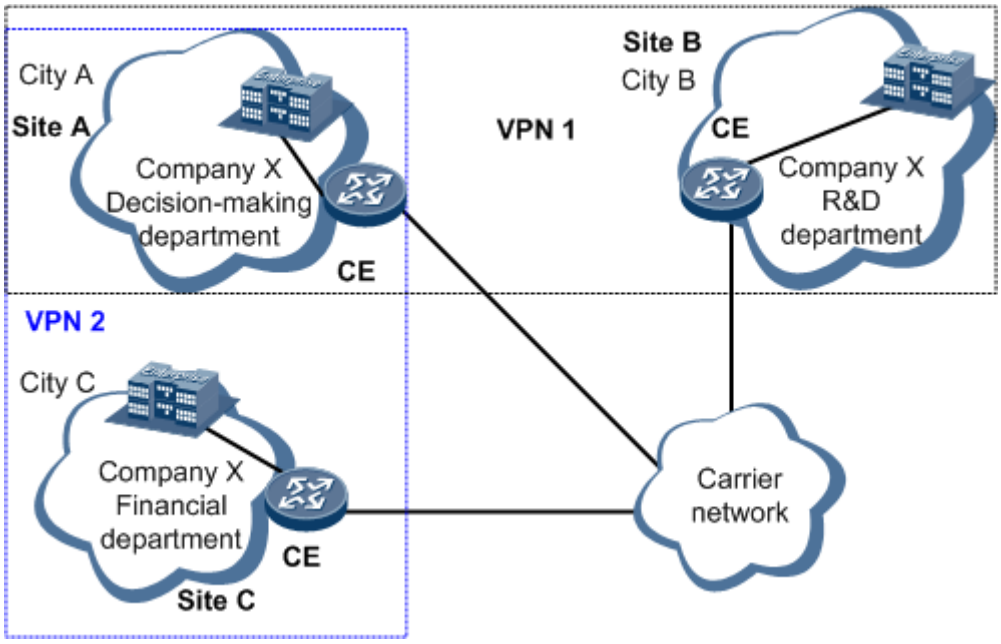
 Copyright © Huawei Technologies Co., Ltd.

As shown in **Figure 9-61**, if the branch network in City B connects to the headquarters network in City A over a leased line instead of a carrier network, the branch network and the headquarters network compose a site.

- The devices at a site may belong to multiple VPNs. In other words, a site may belong to more than one VPN.

  As shown in **Figure 9-62**, the decision-making department of company X in City A (Site A) is allowed to communicate with the R&D department in City B (Site B) and the financial department in City C (Site C). Site B and Site C are not allowed to communicate with each other. In this case, two VPNs, VPN1 and VPN2, can be established, with Site A and Site B belonging to VPN1 and Site A and Site C belonging to VPN2. In this manner, Site A is configured to belong to multiple VPNs.

  Figure 9-62 One site belonging to multiple VPNs



- A site connects to a carrier network through the CE and may contain more than one CE, but a CE belongs only to one site.

  It is recommended that you determine the devices to be used as CEs based on the following principles:

  If the site is a host, use the host as the CE.

  If the site is a subnet, use switches as CEs.

  If the site comprises multiple subnets, use routers as CEs.

  Sites connecting to the same carrier's network can be categorized into different sets based on configured policies. Only sites that belong to the same set can access each other, and this set is a VPN.

 Copyright © Huawei Technologies Co., Ltd.

Address space overlapping

As a private network, a VPN independently manages an address space. Address spaces of different VPNs may overlap. For example, if both VPN1 and VPN2 use addresses on network segment 10.110.10.0/24, address space overlapping occurs.

NOTE

VPNs can use overlapped address spaces in the following situations:

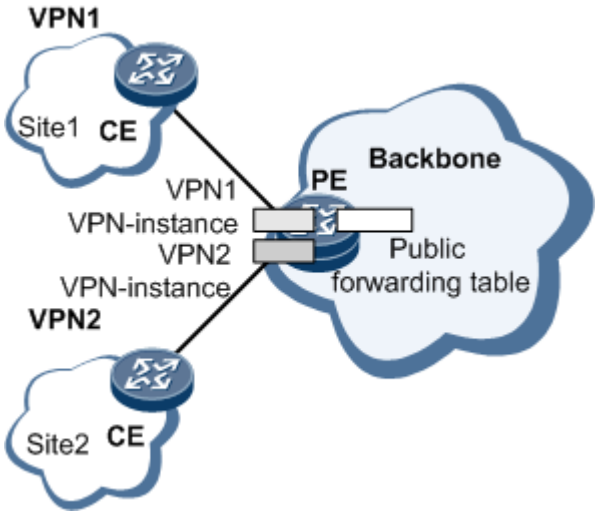
- Two VPNs do not cover the same site.
- Two VPNs cover the same site, but devices at the site and devices using addresses in overlapped address spaces in the VPNs do not access each other.

VPN instance

CEs are user-side devices and need to send only local VPN routes to PEs, irrespective of whether the PEs connect to the public network or other VPNs. PEs are network-side devices, and a PE generally connects to multiple CEs from different VPNs. A PE may receive routes from different VPNs. Because address spaces used by different VPNs may overlap, routes sent from different VPNs may carry the same destination address. If a PE maintains only one routing and forwarding table, this table will accept only one of the routes from different VPNs but with the same destination address. To prevent this problem, the VPN technology uses VPN instances.

A VPN instance is also called a VPN routing and forwarding (VRF) table. A PE maintains multiple routing and forwarding tables, including a public routing and forwarding table and one or more VRF tables. In other words, a PE has multiple instances, including a public network instance and one or more VPN instances, as shown in **Figure 9-63**. Each VPN instance maintains routes from the corresponding VPN. The public network instance maintains public network routes. This enables a PE to keep all routes from VPNs, irrespective of whether their address spaces overlap.

Figure 9-63 Schematic diagram of VPN instances



 Copyright © Huawei Technologies Co., Ltd.

The differences between a public routing and forwarding table and a VRF table are as follows:

- A public routing table contains the IPv4 routes of all PEs and Ps. These IPv4 routes are static routes configured on the backbone network or are generated by routing protocols configured on the backbone network.
- A VPN routing table contains the routes of all sites that belong to the corresponding VPN instance. The routes are obtained through exchange of VPN routes between PEs or between CEs and PEs.
- According to route management policies, a public forwarding table contains the minimum forwarding information extracted from the corresponding routing table, whereas a VPN forwarding table contains the minimum forwarding information extracted from the corresponding VPN routing table.

  The VPN instances on a PE are independent of each other. They are also independent of the public routing and forwarding table.

  Each VPN instance can be regarded as a virtual router, which maintains an independent address space and has one or more interfaces connected to the router.

  In RFC 4364 (BGP/MPLS IP VPNs), a VPN instance is called a per-site forwarding table. As the name suggests, one VPN instance corresponds to one site. To be specific, every connection between a CE and a PE corresponds to a VPN instance, but this is not a one-to-one mapping. The VPN instance is manually bound to the PE interface that directly connects to the CE.

  A VPN instance uses an RD to identify an independent address space and uses VPN targets to manage VPN memberships and routing principles of directly connected sites and remote sites.

Relationships between VPNs, sites, and VPN instances

The relationships between VPNs, sites, and VPN instances are as follows:

- A VPN consists of multiple sites. A site may belong to multiple VPNs.
- A site is associated with a VPN instance on a PE. A VPN instance integrates the VPN member relationships and routing principles of its associated sites. Multiple sites form a VPN based on VPN instance rules.

RD and VPN-IPv4 address

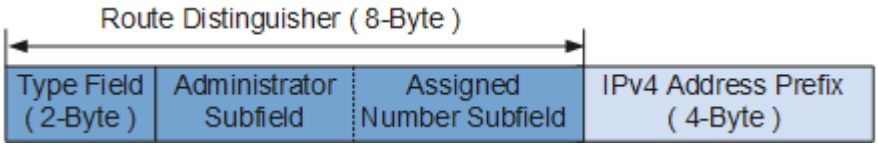
Traditional BGP cannot process the routes of VPNs with overlapped address spaces. Assume that VPN1 and VPN2 use addresses on the network segment 10.110.10.0/24, and each of them advertises a route destined for this network segment. The local PE identifies the two VPN routes based on VPN instances and sends them to the remote PE. Because routes from different VPNs cannot work in load-balancing mode, the remote PE adds only one of the two routes to its VRF table based on BGP route selection rules.

This is because BGP cannot distinguish VPN routes with the same IP address prefix. To solve this problem, BGP/MPLS IP VPN uses the VPN-IPv4 address family.

A VPN-IPv4 address consists of 12 bytes. The first eight bytes represent the RD and the last four bytes represent the IPv4 address prefix, as shown in **Figure 9-64**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-64 VPN-IPv4 address



RDs are used to distinguish IPv4 prefixes using the same address space. The format of RDs enables carriers to allocate RDs independently. An RD, however, must be unique on the entire network to ensure correct routing if CEs are dual-homed to PEs. IPv4 addresses with RDs are called VPN-IPv4 addresses. After receiving IPv4 routes from a CE, a PE converts the routes to globally unique VPN-IPv4 routes and advertises the routes on the public network.

VPN target

The VPN target, also called the route target (RT), is a 32-bit BGP extended community attribute. BGP/MPLS IP VPN uses VPN targets to control the advertisement of VPN routing information.

A VPN instance is associated with one or more VPN targets, which are of the following types:

- Export VPN target: After learning an IPv4 route from a directly connected site, a PE converts the route to a VPN-IPv4 route and sets the export VPN target for the route. As an extended community attribute, the export VPN target is advertised with the route.
- Import VPN target: After receiving a VPN-IPv4 route advertised by another PE, the local PE checks the export VPN target of the route. If the export VPN target is identical with the import VPN target of a VPN instance on the PE, the PE adds the route to the VPN instance.

The VPN target defines the sites that can receive a VPN route, and the sites from which the PE can receive routes.

After receiving a route from a directly connected CE, a PE sets the export VPN targets of the route. The PE then uses BGP to advertise the route with export VPN targets to related PEs. After receiving the route, the related PEs compare the export VPN targets with the import VPN targets of all their VPN instances. If an export VPN target is identical with an import VPN target, the route is added to the corresponding VPN instance.

The reasons for using VPN targets instead of RDs as the extended community attributes are as follows:

- A VPN-IPv4 route has only one RD, but can be associated with multiple VPN targets. With multiple extended community attributes, BGP can greatly improve network flexibility and expansibility.
- VPN targets are used to control route advertisement between different VPNs on a PE. After being configured with matching VPN targets, different VPN instances on a PE can import routes from each other.

 Copyright © Huawei Technologies Co., Ltd.

On a PE, different VPNs have different RDs, but the extended community attributes allowed by BGP are limited. Using RDs for route importing limits network expansibility.

On a BGP/MPLS IP VPN, VPN targets can be used to control exchange of VPN routes between sites. Export VPN targets and import VPN targets are independent of each other and can be configured with multiple values, ensuring flexible VPN access control and diversified VPN networking modes.

Multiprotocol Border Gateway Protocol (MP-BGP)

Traditional BGP-4 defined in RFC 1771 can manage IPv4 routes, but not the routes of VPNs with overlapped address spaces.

To correctly process VPN routes, VPNs use MP-BGP defined in RFC 4760 (Multiprotocol Extensions for BGP-4). MP-BGP supports multiple network layer protocols. Network layer protocol information is contained in the Network Layer Reachability Information (NLRI) field and the Next Hop field of an MP-BGP Update message.

MP-BGP uses the address family to differentiate network layer protocols. An address family can be a traditional IPv4 address family or any other address family such as a VPN-IPv4 address family or an IPv6 address family. For the values of address families, see RFC 3232 (Assigned Numbers).

9.5.5 BGP/MPLS IP VPN Principles

Route Advertisement on a Basic BGP/MPLS IP VPN

On a basic BGP/MPLS IP VPN, CEs and PEs are responsible for advertising VPN routes, whereas Ps only need to maintain backbone network routes without knowing VPN routing information. Generally, a PE maintains the routes of VPNs that the PE accesses, rather than all VPN routes. VPN route advertisement consists of the following phases: Route advertisement from the local CE to the ingress PE Route advertisement from the ingress PE to the egress PE Route advertisement from the egress PE to the remote CE After the process of route advertisement is complete, the local and remote CEs can set up reachable routes, and VPN routing information can be advertised on the backbone network. The following describes the three phases of route advertisement in detail:

1. Route advertisement from the local CE to the ingress PE

   After the peer relationship is set up between a CE and the directly connected PE, the CE advertises the local IPv4 routes to the PE. The CE can communicate with the PE over static routes or routes established using Routing Information Protocol (RIP), Open Shortest Path First (OSPF), Intermediate System to Intermediate System (IS-IS), or BGP. Routes advertised by the CE to the PE are standard IPv4 routes, regardless of which routing protocol is used.

   VPN instances on a PE are isolated from each other and independent of the public routing and forwarding table, so as to prevent problems caused by **address space overlapping**. After learning routes from CEs, a PE decides to which routing and forwarding table the routes should be installed. Common static routes and routing protocols do not have this capability, and manual configuration is required.

 Copyright © Huawei Technologies Co., Ltd.

2. Route advertisement from the ingress PE to the egress PE

   Route advertisement from the ingress PE to the egress PE consists of the following phases:

   - After learning VPN routes from a CE, a PE stores these routes in corresponding VRFs and adds **RDs** to these standard IPv4 routes. The VPN-IPv4 routes are then generated.
   - The ingress PE advertises VPN-IPv4 routes to the egress PE by sending **MP-BGP** Update messages. The MP-BGP Update messages also contain **VPN targets** and MPLS labels.

   Before the next-hop PE receives the VPN-IPv4 routes, the routes are first filtered by BGP routing policies, including the export policy configured on the VPN instance and the peer export policy.

   After these routes arrive at the egress PE, if they match the BGP peer import policy and their next hops are reachable or they can be iterated, the egress PE performs local route crossing and filters these routes based on a VRF import policy. The egress PE then decides which routes are to be added to its VPN routing tables. Routes received from other PEs are added to a VPN routing table based on **VPN targets**. The egress PE stores the following information for subsequent packet forwarding:

   - Values of MPLS labels contained in MP-BGP Update messages
   - Tunnel IDs generated after tunnel iteration

3. Route advertisement from the egress PE to the remote CE

   A remote CE can learn VPN routes from an egress PE over static routes or routes established using RIP, OSPF, IS-IS, or BGP. Route advertisement from the egress PE to a remote CE is similar to that from a local CE to the ingress PE. The details are not described here. Note that the routes advertised by the egress PE to the remote CE are standard IPv4 routes.

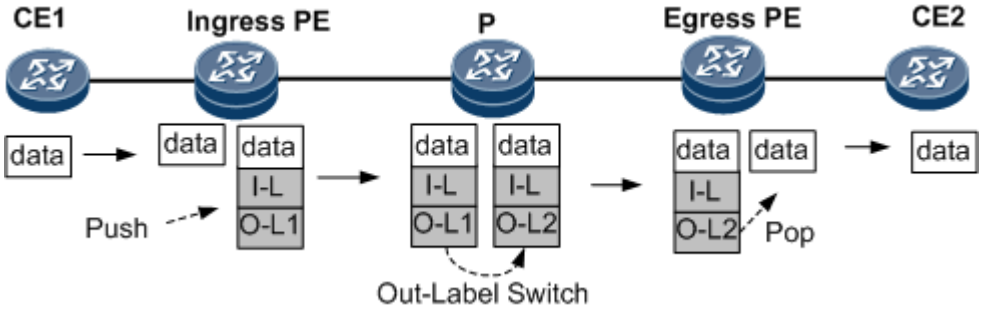
After a PE receives routes of different VPNs from a local CE, if the next hops of these routes are reachable or these routes can be iterated, the PE matches the export VPN targets of these routes with the import VPN targets of its local VPN instances. This process is called local route crossing. During local route crossing, the PE filters these routes based on a VRF import policy and modifies the attributes of eligible routes.

Packet Forwarding on a BGP/MPLS IP VPN

On a BGP/MPLS IP VPN backbone network, a P does not know VPN routing information. VPN packets are forwarded between PEs over tunnels. **Figure 9-65** shows an example of packet forwarding on a BGP/MPLS IP VPN. A packet is transmitted from CE1 to CE2. I-L indicates an inner label, and O-L indicates an outer label. The outer label directs the packet to the BGP next hop, and the inner label identifies the outbound interface for the packet or the VPN to which the packet belongs.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-65 Forwarding of a VPN packet from CE1 to CE2



The forwarding process is as follows:

1. CE1 sends a VPN packet to the ingress PE.
2. After receiving the packet from an interface bound to a VPN instance, the ingress PE performs the following steps:
   - Searches the corresponding VPN forwarding table based on the RD of the bound VPN instance.
   - Matches the destination IPv4 address with forwarding entries and searches for the corresponding tunnel ID.
   - Adds an I-L to the packet and finds the tunnel to be used based on the tunnel ID.
   - Adds an outer label to the packet and sends the packet over the tunnel. In this example, the tunnel is an LSP, and the outer label is an MPLS label.
   - Transmits the double-tagged packet over the backbone network. Each P on the forwarding path swaps the outer label of the packet.
3. After receiving the packet, the egress PE removes the outer label of the packet.

   NOTE

   In this example, the final outer label of the packet is O-L2. If PHP is configured, O-L2 is removed on the penultimate hop, and the egress PE receives a packet with the inner label only.
4. The egress PE removes the inner label residing at the bottom of the label stack.
5. The egress PE sends the packet from the corresponding outbound interface to CE2. After its labels are removed, the packet becomes a pure IP packet.

In this manner, the packet is sent from CE1 to CE2. CE2 forwards the packet to the destination in the way it sends other IP packets.

9.5.6 Hub & Spoke

The Hub & Spoke networking can be used to enable an access control device on a VPN to control the mutual access of other users. The site where the access control device locates is called a Hub site, and other sites are called Spoke sites. At the Hub site, a device that accesses the VPN backbone network is called a Hub-CE; at a Spoke site, a device that accesses the VPN backbone network is called a Spoke-

 Copyright © Huawei Technologies Co., Ltd.

CE. On the VPN backbone network, a device that accesses the Hub site is called a Hub-PE; a device that accesses a Spoke site is called a Spoke-PE.

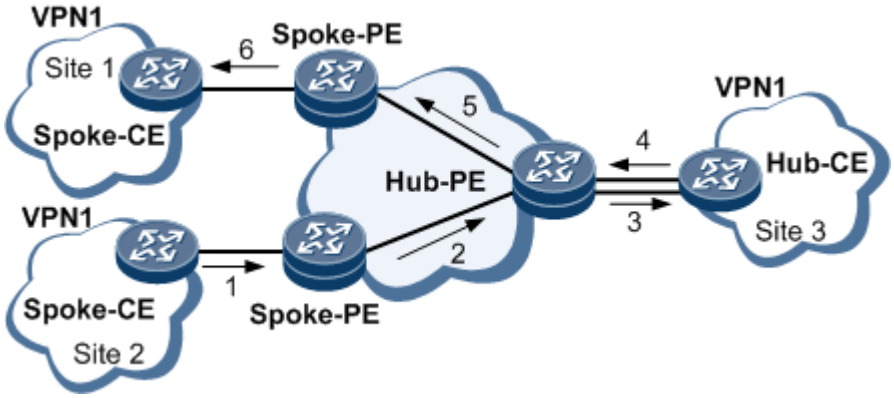
A Spoke site advertises routes to the Hub site, and the Hub site then advertises the routes to other Spoke sites. No direct route exists between the Spoke sites. The Hub site controls the communication between the Spoke sites.

In the Hub & Spoke networking model, two VPN targets are configured to stand for Hub and Spoke respectively.

The configuration of a VPN target on a PE must comply with the following rules:

- The export target and the import target of the Spoke-PE at a Spoke site are Spoke and Hub respectively. The import target of a Spoke-PE is different from the export targets of other Spoke-PEs.
- A Hub-PE requires two interfaces or sub-interfaces. One interface or sub-interface receives routes from Spoke-PEs, and the import target of the VPN instance on the interface is Spoke. The other interface or sub-interface advertises the routes to Spoke-PEs, and the export target of the VPN instance on the interface is Hub.

Figure 9-66 Route advertisement from Site 2 to Site 1 in Hub & Spoke networking



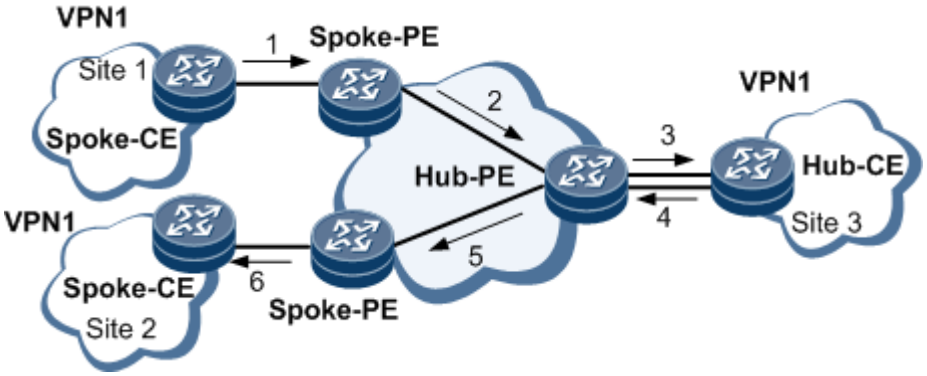
As shown in **Figure 9-66**, the communication between Spoke sites is controlled by the Hub site. The lines with arrowheads show the process of advertising a route from Site 2 to Site 1.

- The Hub-PE can receive the VPN-IPv4 routes advertised by all the Spoke-PEs.
- All the Spoke-PEs can receive the VPN-IPv4 routes advertised by the Hub-PE.
- The Hub-PE advertises the routes learned from the Spoke-PEs to the Hub-CE, and advertises the routes learned from the Hub-CE to all the Spoke-PEs. The Spoke sites can access each other through the Hub site.
- The import target of a Spoke-PE is different from the export targets of other Spoke-PEs. Two Spoke-PEs cannot directly advertise VPN-IPv4 routes to each other. As a result, the Spoke sites cannot access each other.

The transmission path between Site 1 and Site 2 is shown in **Figure 9-67**. The lines with arrowheads indicate the path from Site 2 to Site 1.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-67 Path of data transmission from Site 1 to Site 2



Networking Description

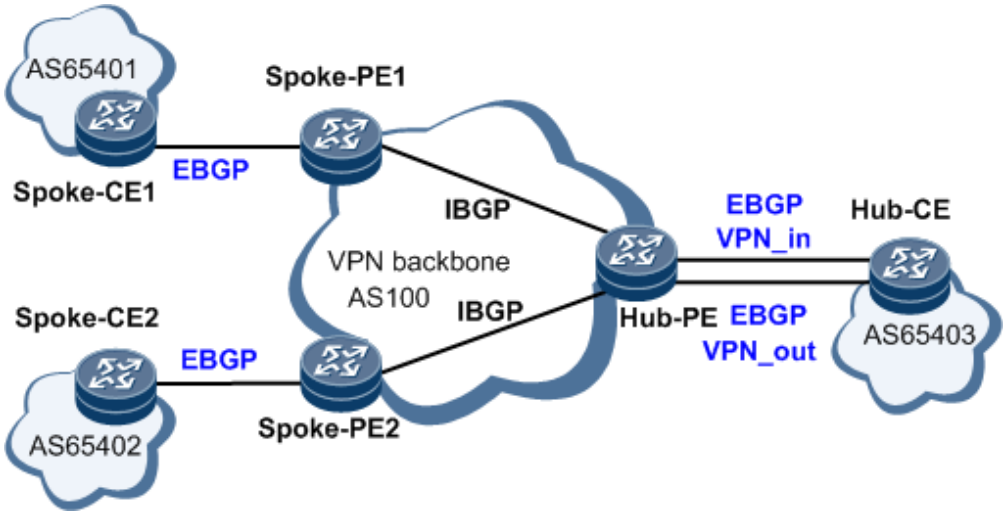
Hub & Spoke networking schemes include:

- External Border Gateway Protocol (EBGP) running between the Hub-CE and Hub-PE, and between Spoke-PEs and Spoke-CEs
- IGP running between the Hub-CE and Hub-PE, and between Spoke-PEs and Spoke-CEs
- EBGP running between the Hub-CE and Hub-PE, and IGP running between Spoke-PEs and Spoke-CEs

The following describes these networking schemes in detail:

- EBGP running between the Hub-CE and Hub-PE, and between Spoke-PEs and Spoke-CEs

  Figure 9-68 EBGP running between the Hub-CE and Hub-PE, and between Spoke-PEs and Spoke-CEs



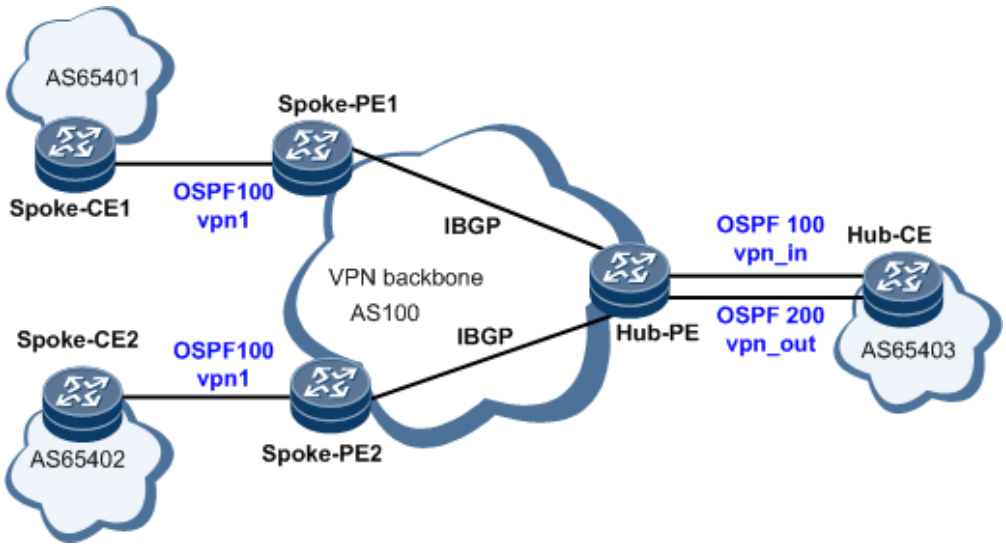
  As shown in **Figure 9-68**, the routing information advertised by a Spoke-CE is forwarded to the Hub-CE before being transmitted to other Spoke-PEs. If

 Copyright © Huawei Technologies Co., Ltd.

EBGP runs between the Hub-PE and Hub-CE, the Hub-PE performs the AS-Loop check on the route. If the Hub-PE detects its own AS number in the route, it discards the route. In this case, to implement the Hub & Spoke networking, the Hub-PE must be configured to permit the existence of repeated local AS numbers.

- IGP running between the Hub-CE and Hub-PE, and between Spoke-PEs and Spoke-CEs

  Figure 9-69 IGP running between the Hub-CE and Hub-PE, and between Spoke-PEs and Spoke-CEs

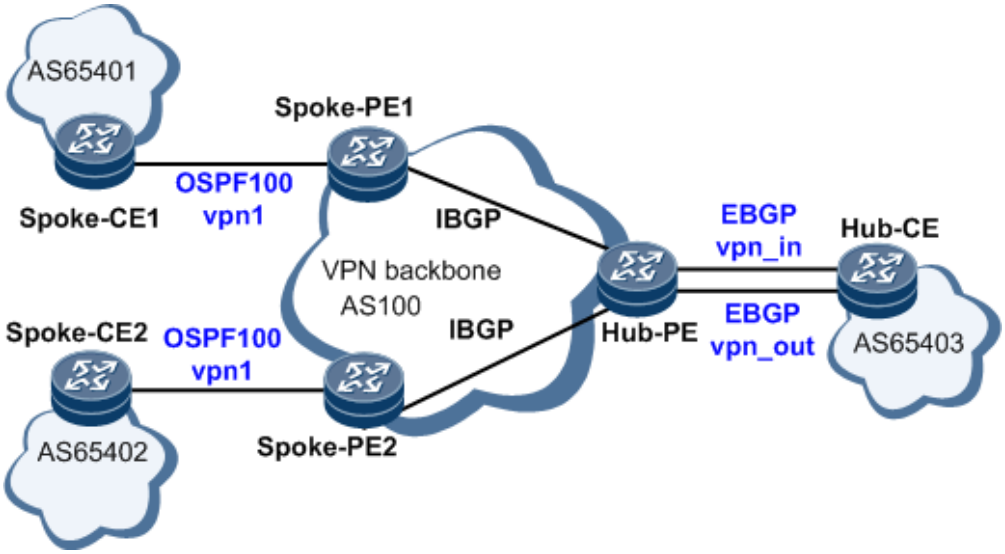
  

  Because all PEs and CEs exchange routing information through IGP and IGP routes do not contain the AS_Path attribute, the AS_Path field of BGP VPNv4 routes is null.

- EBGP running between the Hub-CE and Hub-PE, and IGP running between Spoke-PEs and Spoke-CEs

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-70 EBGP running between the Hub-CE and Hub-PE, and IGP running between Spoke-PEs and Spoke-CEs



The networking topology is similar to that shown in **Figure 9-68**. The AS_Path attribute of the route forwarded by the Hub-CE to the Hub-PE contains the AS number of the Hub-PE. Therefore, the Hub-PE must be configured to permit the existence of repeated local AS numbers.

9.5.7 MCE

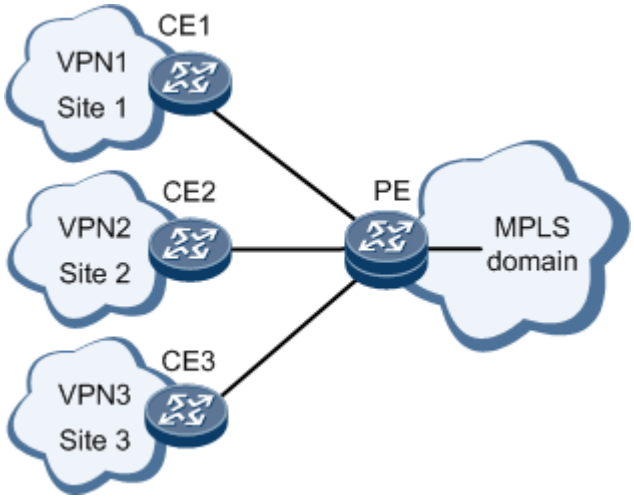
The multi-VPN-instance customer edge (MCE) technology provides logically independent VPN instances and address spaces on a CE, allowing multiple VPN users to share the same CE. The MCE technology provides an economical and easy-to-use solution to solve problems concerned with VPN service isolation and security.

Background

VPN services are becoming increasingly refined, and the demand for VPN service security is growing. Carriers must isolate different types of VPN services on networks to meet this demand. As shown in **Figure 9-71**, the traditional BGP/MPLS IP VPN technology isolates VPN services by deploying one CE for each VPN, bringing in high costs and complicated network deployment. If multiple VPNs use the same CE to access upper-layer devices, these VPNs will share the same routing and forwarding table, and data security for these VPNs cannot be ensured. The MCE technology addresses conflicts between network costs and data security problems caused by multiple VPNs sharing the same CE.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-71 Networking diagram for VPN service isolation using BGP/MPLS IP VPN

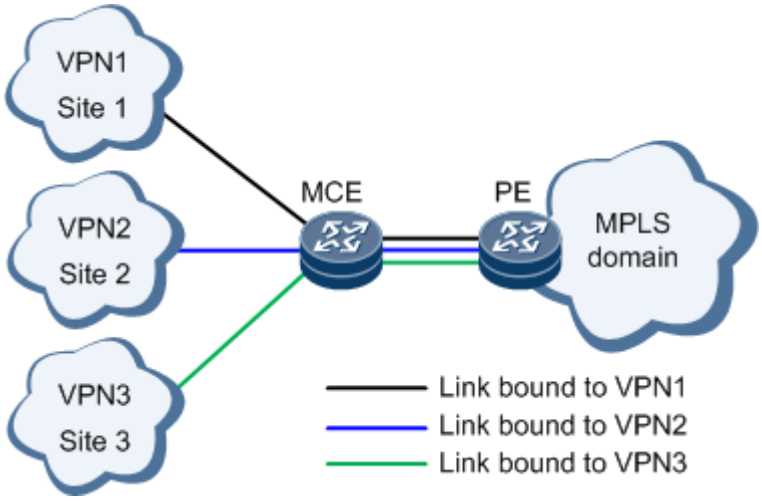


Implementation

The MCE technology creates a VPN instance for each VPN service to be isolated. Each VPN uses an independent routing protocol to communicate with the MCE to which these VPNs are connected. A VPN instance is bound to each link between the MCE and the PE to which the MCE is bound. As a result, an independent channel is established for each VPN service, and different VPN services are isolated.

As shown in **Figure 9-72**, three VPN instances are configured on the MCE: VPN1, VPN2, and VPN3. To be specific, three independent VPN routing and forwarding tables are created on the MCE. VPN1 is bound to the link between the MCE and Site1 and a link between the MCE and PE, VPN2 is bound to the link between the MCE and Site2 and a link between the MCE and PE, and VPN3 is bound to the link between the MCE and Site3 and a link between the MCE and PE. These configurations allow VPN services to be isolated using only one MCE.

Figure 9-72 MCE networking



 Copyright © Huawei Technologies Co., Ltd.

Benefits

The MCE technology enables CEs to provide PE functions. MCEs avoid the practice of deploying one CE for each VPN while isolating VPN services, significantly reducing maintenance costs and expenditure on devices.

9.5.8 BGP/MPLS IP VPN Label Distribution Modes

Background

On a BGP/MPLS IP VPN, a private network label (MPLS label) is assigned for each VPN route by default. If a large number of VPN routes exist, MPLS label resources may be insufficient. Other label distribution modes can be used to save MPLS label resources. **Table 9-14** compares the label distribution modes which are currently available.

Table 9-14 Comparison among label distribution modes

Label Distribution Mode	Description	Applicable Networking	Configuration Position
One label per instance	All the VPN routes from a VPN instance are assigned the same private network label.	All types of BGP/MPLS IP VPNs	Devices on which VPN instances are configured
One label per next hop	All the VPN routes with the same next hop and outgoing label are assigned the same private network label.	Inter-AS VPN Option B	ASBRs

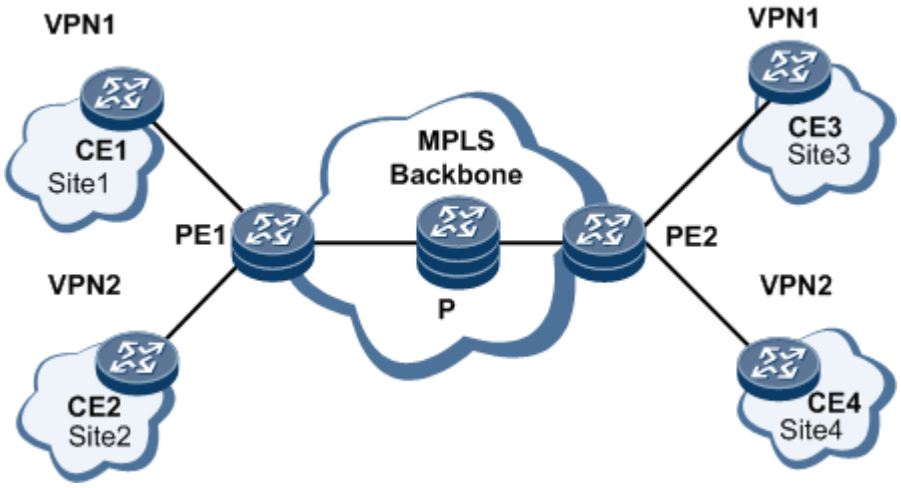
Implementation

One Label per Instance

After one-label-per-instance label distribution is configured in the IPv4 or IPv6 address family of a VPN instance, all the VPN routes from the corresponding address family of the VPN instance share the same private network label. On the network shown in **Figure 9-73**, PE1 has two VPN instances: VPN1 and VPN2. If PE1 receives 10,000 VPN routes from VPN1 and VPN2, respectively, PE1 assigns 20,000 private network labels to these routes by default. After one-label-per-instance label distribution is configured, PE1 assigns only one label to the 10,000 VPN routes received from VPN1 and one label to the 10,000 VPN routes received from VPN2.

 Copyright © Huawei Technologies Co., Ltd.

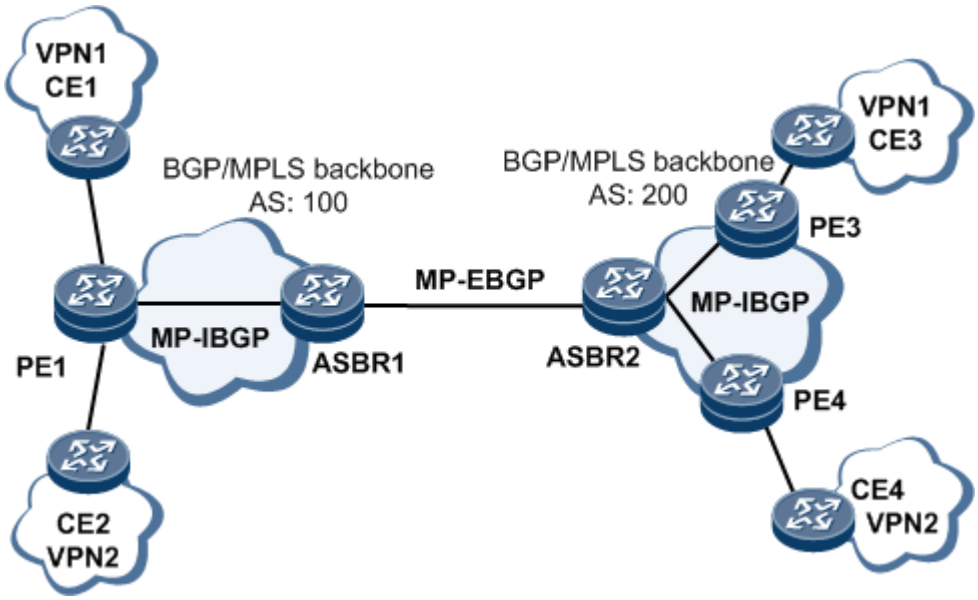
Figure 9-73 One label per instance



One label per next hop

By default, an ASBR assigns a private network label to each VPN route. You can also configure an ASBR to distribute labels based on next hops. After one-label-per-next-hop label distribution is configured on an ASBR, the ASBR re-advertises an MP-BGP Update packet to its peers. The MP-BGP Update packet carries VPNv4 routes and their labels that are re-assigned based on next hops. After a peer receives the MP-BGP Update packet, the peer updates its local label forwarding table and re-establishes LSPs. After the ASBR and its peers update their label forwarding tables, packets are forwarded based on the updated label forwarding tables.

Figure 9-74 One label per next hop



 Copyright © Huawei Technologies Co., Ltd.

In the inter-AS VPN Option B scenario shown in **Figure 9-74**, two VPN instances named VPN1 and VPN2 are configured on PE1 that uses one-label-per-instance label distribution. If one-label-per-next-hop label distribution is not configured on ASBR1, ASBR1 assigns 20,000 labels to the 20,000 VPN routes imported from CE1 and CE2 over PE1 before advertising these routes to ASRB2. After one-label-per-next-hop label distribution is configured on ASBR1, ASBR1 only needs to assign one label to VPN routes with the same next hop and outgoing label. In other words, ASBR1 assigns only two labels to the 20,000 routes imported from CE1 and CE2 over PE1 before advertising these routes to ASRB2.

NOTE

One-label-per-route label distribution and one-label-per-next-hop label distribution can be flexibly switched to each other. During label distribution mode switching, service packets are lost for a short period due to the update of label forwarding tables.

In an inter-AS VPN Option B scenario, be sure to configure one-label-per-instance label distribution on PEs if you configure one-label-per-next-hop label distribution on ASBRs.

Benefits

Using an appropriate label distribute mode saves label resources.

9.5.9 BGP Site of Original (SoO)

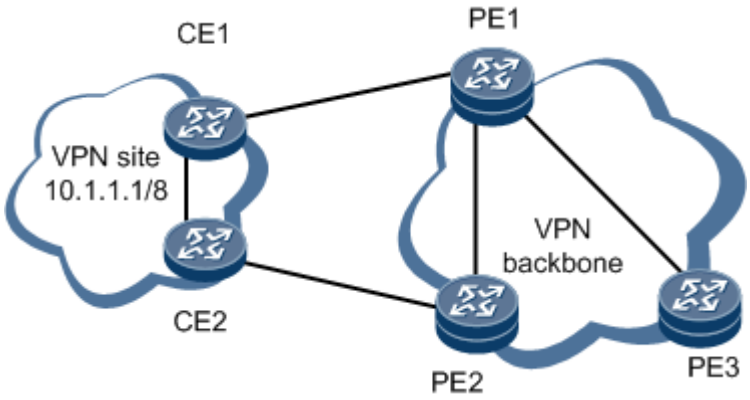
If multiple CEs in a VPN site access different PEs, VPN routes sent from CEs to PEs may return to this VPN site after traveling across the backbone network. This may cause routing loops in the VPN site.

After the SoO attribute is configured on a PE, the PE adds the SoO attribute to the route sent from a CE and then advertises the route to other PE peers. Before advertising the VPN route to the connected CE, the PE peers check the SoO attribute carried in the VPN route. If the PE peers find that this SoO attribute is the same as the locally configured SoO attribute, the PE peers do not advertise this VPN route to the connected CE.

On the network shown in **Figure 9-75**, CE1 and CE2 belong to the same VPN site and can advertise routes to each other. CE1 advertises the route destined for 10.1.1.1/8 in the VPN site to PE1, and PE1 advertises the route to PE2 by using Multiprotocol Internal Border Gateway Protocol (MP-IBGP). PE2 then advertises the route to CE2 by using BGP. As a result, the route returns to the original VPN site from which the route is advertised, which may cause a routing loop in the VPN site.

 Copyright © Huawei Technologies Co., Ltd.

Figure 9-75 Networking diagram for BGP SoO application



To avoid routing loops in a VPN site, you can configure an SoO attribute on PE1 for CE1. The SoO attribute identifies the site where the CE1 resides. The routes advertised by CE1 to PE1 then carry this SoO attribute, and PE1 advertises the routes with the SoO attribute to other PEs across the backbone network. Before advertising the received routes to its peer CE2, PE2 checks whether the routes carry the SoO attribute specified for the site where CE2 resides. If a route carries this SoO attribute, this route is advertised from the site where CE2 resides. PE2 then refuses to advertise such a route to CE2, avoiding routing loops in the site.

9.5.10 Terms, Acronyms, and Abbreviations for BGP/MPLS IP VPN

Terms

Term	Definition
CE	Customer edge equipment that is directly connected to the service provider. On an MPLS VPN, a CE can be a router, switch, or even a host.
Address space	An address realm managed by a VPN.
L2TP	A Layer 2 tunneling protocol that is drafted by the IETF and involves the participation of companies such as Microsoft. L2TP combines the advantages of both PPTP and L2F.
MP-BGP	A multi-protocol extension of BGP-4. MP-BGP supports multiple network layer protocols and identifies the protocols based on address families. MP-BGP transmits VPN composition information and VPN-IPv4 routes between PEs.
P	A backbone router that is located on the service provider network. A P is indirectly connected to CEs. Ps only need to possess basic MPLS forwarding capabilities and do not maintain information about a VPN.

 Copyright © Huawei Technologies Co., Ltd.

Term	Definition
PE	A router that is located on the backbone network of the MPLS VPN structure. A PE is responsible for VPN user management, establishment of LSPs between PEs, and exchange of routing information between sites of the same VPN. During the process, a PE performs the mapping and forwarding of packets from the private network to the public channel and then from the public channel to the private network. A PE can be a UPE, an SPE, or an NPE.
RD	A route identifier, which is an 8-byte field in a VPN IPv4 address. An RD and a 4-byte IPv4 address prefix construct a VPN IPv4 address, which is used to differentiate the IPv4 prefixes using the same address space.
Site	A group of IP systems with IP connectivity, which can be achieved independent of SP networks.
Tunnel	A channel through which a packet switching network transmits service traffic between PEs. In VPN, a tunnel is an information transmission channel between two entities. The tunnel provides security for transparent transmission of VPN information. In most cases, a tunnel is an MPLS tunnel.
Tunnel iteration	A process in which a route is iterated to a tunnel.
Tunnel key	A group of information, including the token, slot number of an outgoing interface, tunnel type, and location method.
VPN	A recently-developed technology that implements the private network over a public network. It is a network that only logically exists.
VPN instance	An entity that is set up and maintained by PEs for directly-connected sites. Each site has its VPN instance on a PE. A VPN instance is also called a VRF table. A PE has multiple forwarding tables, including a public-network routing table and one or more VRF tables.
VPN target	A BGP extended community attribute that is also called route target. In BGP/MPLS IP VPN, VPN targets are used to control VPN routing information. The VPN target attribute defines that a VPN IPv4 route can be received by which sites and a PE device can receive routes from which site.

Acronyms and Abbreviations

Acronym and Abbreviation	Full Name
AMB	active main board

 Copyright © Huawei Technologies Co., Ltd.

Acronym and Abbreviation	Full Name
AS	autonomous system
ASG	access service gateway
BSC	base station controller
BTS	base transceiver station
DCN	data communication network
EBGP	External Border Gateway Protocol
EGP	Exterior Gateway Protocol
FMC	fixed-mobile convergence
GR	graceful restart
HA	high availability
HoPE	hierarchy of PE
IBGP	Internal Border Gateway Protocol
IS-IS	Intermediate System to Intermediate System
L3VPN	Layer 3 virtual private network
LSDB	link state database
LTE	Long Term Evolution
MCE	multi-VPN-instance customer edge
MME	mobility management entity
MBGP	Multiprotocol Border Gateway Protocol
MP-EBGP	Multiprotocol External Border Gateway Protocol
MP-IBGP	Multiprotocol Internal Border Gateway Protocol
NLRI	Network Layer Reachability Information
NSF	non-stop forwarding
NSR	non-stop routing
OSPF	Open Shortest Path First
PSTN	public switched telephone network
RIP	Routing Information Protocol
RIPng	Routing Information Protocol next generation
RM	routing management

 Copyright © Huawei Technologies Co., Ltd.

Acronym and Abbreviation	Full Name
RNC	radio network controller
RP	route processor
RR	route reflector
RSG	radio service gateway
RSVP	Resource Reservation Protocol
RT	route target
SP	service provider
SMB	standby main board
TCP	Transport Control Protocol
VRF	VPN routing and forwarding

9.6 GRE

9.6.1 Introduction to GRE

Definition

Generic Routing Encapsulation (GRE) is a common routing encapsulation protocol that can encapsulate network layer protocols, such as IPX, ATM, IPv6/IPv4, and AppleTalk, inside virtual point-to-point links over an Internet Protocol network.

GRE provides a transmission channel in which one network layer protocol is encapsulated in another network layer protocol. This channel is called a GRE tunnel.

Feature Value

- Allows one network layer protocol to be encapsulated in another network layer protocol to traverse heterogeneous networks.
- Provides a workaround for networks that support a limited number of hops.
- Sets up a VPN across two physically separate networks.
- Carries IPv4 packets on private or public networks.
- Lowers requirements on device performance and sets up tunnels between devices that do not support Multiprotocol Label Switching (MPLS).

 Copyright © Huawei Technologies Co., Ltd.

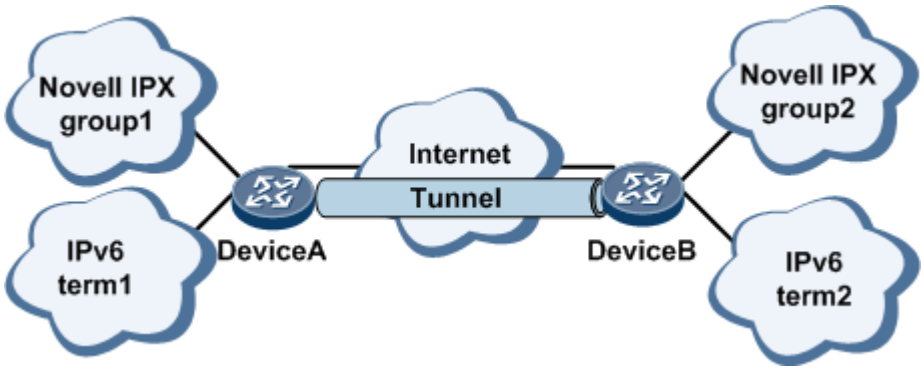
9.6.2 GRE Basic Principles

Background

A single network protocol, such as IPv4, is used to transmit packets on a backbone network, whereas other protocols, such as Internetwork Packet Exchange (IPX), Asynchronous Transfer Mode (ATM), IPv6, may be used on non-backbone networks. Because the backbone and non-backbone networks use different protocols, packets cannot be transmitted between the non-backbone networks over the backbone network. Generic Routing Encapsulation (GRE) resolves this issue by encapsulating one network protocol inside another.

On the network shown in **Figure 9-76**, groups 1 and 2 are the non-backbone networks running Novell IPX, and teams 1 and 2 are the non-backbone networks running IPv6. The backbone network is an IPv4 network. To transmit packets between groups 1 and 2 and between teams 1 and 2 over the backbone network, GRE is used to establish a tunnel between DeviceA and DeviceB. When DeviceA receives a packet from group 1 or team 1, DeviceA encapsulates the packet into a GRE packet. The GRE packet is then encapsulated into an IPv4 packet for forwarding.

Figure 9-76 GRE networking

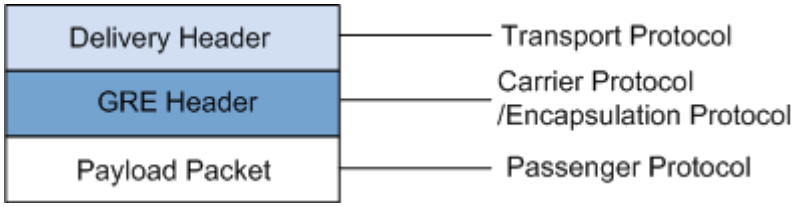


Related Concepts

- GRE packet format

  After receiving a packet of network layer protocol, the system adds a GRE header to the packet and encapsulates the packet into another protocol, such as IP. Then, the IP protocol forwards the packet. **Figure 9-77** shows the format of a GRE packet.

  Figure 9-77 Format of a GRE packet

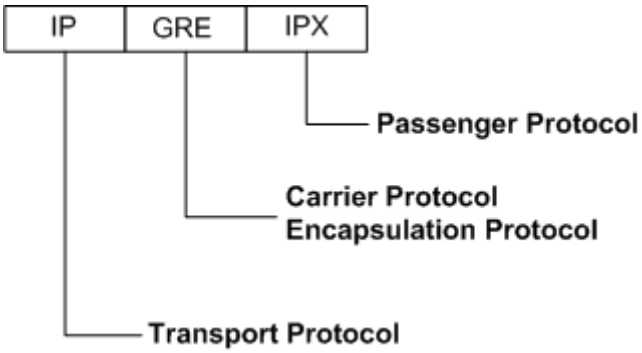


 Copyright © Huawei Technologies Co., Ltd.

- Payload: is received by the system and needs to be encapsulated and routed.
- Passenger protocol: is used by the packet before encapsulation.
- Encapsulation protocol: is used to encapsulate the passenger protocol. It is also called the carrier protocol.
- Transport or delivery protocol: is responsible for forwarding the encapsulated packets.

The following shows the format of an IPX packet encapsulated for transmission over an IP tunnel.

Figure 9-78 Format of an IPX packet transmitted over an IP tunnel



NOTE

For an IPv4 GRE tunnel:

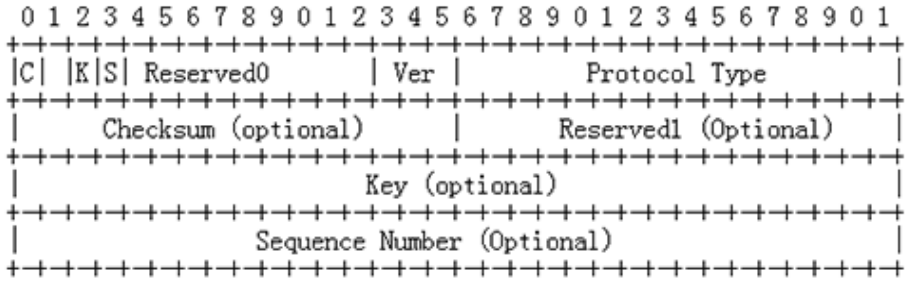
- The packet transmission protocol is IPv4.
- Only IPv4 packets can be encapsulated and routed.

- GRE packet header

  RFC 1701, RFC 2784, and RFC 2890 define the GRE packet format. This feature is implemented according to RFC 2784 and RFC 2890 and is partially compatible with RFC 1701.

  Figure 9-79 shows the format of the GRE header defined in RFC 2890.

Figure 9-79 Format of the GRE packet header defined in RFC 2890



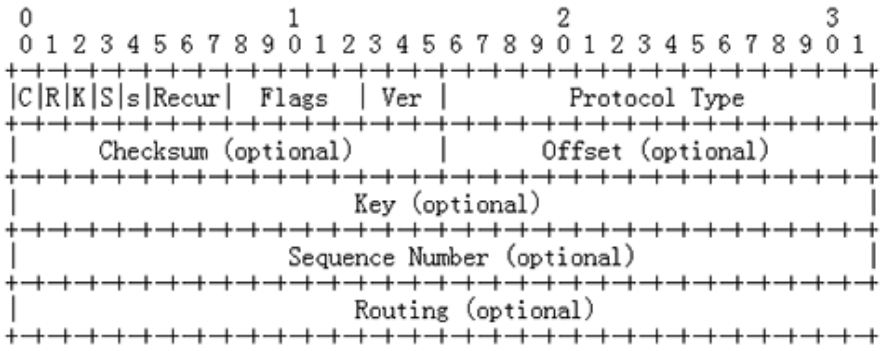
Each field is described as follows:

 Copyright © Huawei Technologies Co., Ltd.

- C: indicates checksum. If its value is 1, the checksum field is inserted into the GRE header. If its value is 0, the GRE header does not contain the checksum field.
- K: indicates keyword. If its value is 1, the key field is inserted into the GRE header. If its value is 0, the GRE header does not contain the key field. This bit in RFC 2890 is used to identify the type of traffic in a GRE tunnel. The meaning of a keyword can be defined by customers. For example, it can be used to identify a virtual private network (VPN) to which the traffic inside a GRE tunnel belongs.
- S: indicates the sequence number. If its value is 1, the sequence number field is inserted into the GRE header. If its value is 0, the GRE header does not contain the sequence number field. If a received packet is out of sequence, it might be discarded.
- Reserved0: indicates the reserved bit, filled with 0.
- Ver: indicates the version field, which must be set to 0. Ver 1 is used for RFC 2637 PPTP.
- Protocol Type: indicates the negotiated passenger protocol type.
- Checksum: indicates the checksum field in the GRE header and payload. This field is filled in when the C bit is 1.
- Key: indicates the keyword. This field is filled in when K bit is 1.

Figure 9-80 shows the format of the GRE header defined in RFC 1701.

Figure 9-80 Format of the GRE packet header defined in RFC 1701



Each field is described as follows:

- R: indicates the routing present bit. If its value is 1, the Routing field is filled in.
- s: indicates the strict source route field. If its value is 1, Routing contains Strict Source Route.
- Recur: Indicates the recursion bit, which is used to indicate the number of GRE packet encapsulation layers. If its value is 0, there is only one-layer GRE encapsulation. Its value is increased by 1 for every GRE encapsulation. If the number of encapsulation layers is greater than 3, the packet is dropped. This field prevents a packet from being encapsulated for infinite times.

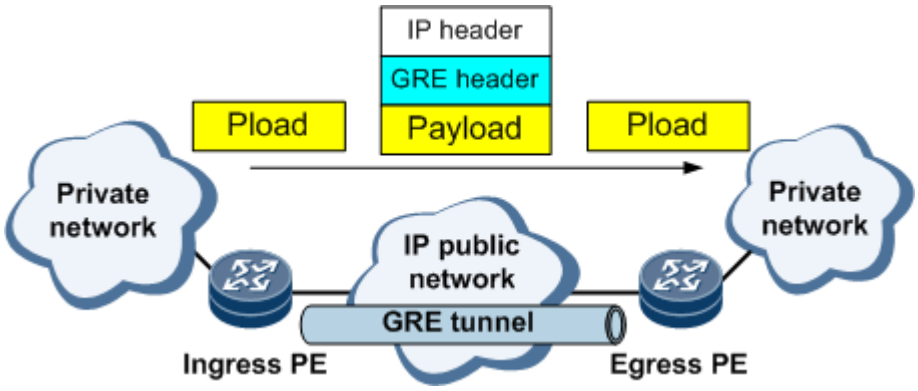
 Copyright © Huawei Technologies Co., Ltd.

- Flags: reserved. It must be set to 0 currently.
- K: indicates the keyword, which is used by the receiver of a tunnel to verify received packets.

Transmission of Packets over a GRE Tunnel

The transmission of packets over a GRE tunnel can be divided into two phases: encapsulation and decapsulation. On the network shown in **Figure 9-81**, a private network packet is encapsulated on the ingress PE and decapsulated on the egress PE.

Figure 9-81 Interconnection of private networks over a GRE tunnel



- Encapsulation

  Upon receiving a private network packet, the ingress PE delivers the packet to the private network protocol module.

  The protocol module checks the destination address field in the packet header, searches the routing table or forwarding table of the private network for the outgoing interface, and determines how to route this packet. If the outgoing interface is a GRE tunnel interface, the protocol module sends the packet to the tunnel module.

  Upon receiving the packet, the tunnel module processes the packet as follows:

  a. Adds a GRE header to the packet according to the protocol type of the passenger protocol and the Key parameter configured for the current GRE tunnel.

  b. Adds a transport protocol header to the packet based on the configuration. For example, if the transport protocol is IP, the source and destination addresses carried in the IP header are the source and destination addresses of the tunnel.

  c. Delivers the packet to the IP module. Based on the destination address in the IP header, the IP module searches the public network routing table for the outgoing interface and sends the packet. The encapsulated packet is then transmitted on the public IP network.

- Decapsulation

  The decapsulation process is opposite to the encapsulation process. Upon receiving a packet, the egress PE analyzes the IP header. After determining

 Copyright © Huawei Technologies Co., Ltd.

that the destination of the packet is itself and the Protocol Type field is 47, which indicates that the protocol is GRE (see RFC 1701 and RFC 2784), the egress PE delivers the packet to the GRE module. The GRE module removes the IP and GRE headers and learns from the Protocol Type field in the GRE header that the passenger protocol is the protocol running on the private network. The GRE module then delivers the packet to the module corresponding to this protocol, which forwards the packet.

9.6.3 GRE Principles

9.6.3.1 Keepalive Detection

Principles

GRE does not support link status detection. If the remote interface is unreachable, the tunnel cannot immediately close the tunnel connection. As a result, the source continues forwarding packets to the peer.

However, link status detection can be provided by the keepalive function, which can be enabled on tunnel interfaces to detect the status of the remote interface.

Implementation

After the keepalive function is enabled, the source GRE tunnel periodically sends probe packets to the peer. If the peer is reachable, the source tunnel interface can receive a reply. The details are as follows:

1. After the keepalive function is enabled on the source GRE tunnel interface, the interface starts a timer, periodically sends probe packets, and starts a probe packet counter, which increases by 1 each time a probe packet is sent.
2. The peer sends a reply to the source after receiving a probe packet.
3. If the source receives a reply before the counter reaches the preset value, the source considers the peer reachable and resets the counter. Otherwise, the source considers the peer unreachable and closes the tunnel connection.

NOTE

If the keepalive function is enabled on a tunnel interface, the remote tunnel interface sends a reply upon receiving a probe packet,

regardless of whether the keepalive function is enabled.

Benefits

The keepalive function prevents data loss when the peer becomes unreachable, ensuring data transmission reliability.

9.6.3.2 GRE Security Mechanism

 Copyright © Huawei Technologies Co., Ltd.

Remote GRE Header Querying

The trace-counter function can capture packets from the remote device to check the generic routing encapsulation (GRE) header and verify whether the packet information is correct during packet transmission and encapsulation.

- **GRE Keyword Authentication**

  GRE supports keyword authentication, a security mechanism defined in RFC 1701 for authenticating tunnel interfaces. This security mechanism prevents tunnel interfaces from incorrectly identifying and receiving packets from other routers.

  As defined in RFC 1701, if the K bit in the GRE header is set to 1, the Key field is inserted to the GRE header, and both the receiver and sender perform key authentication.

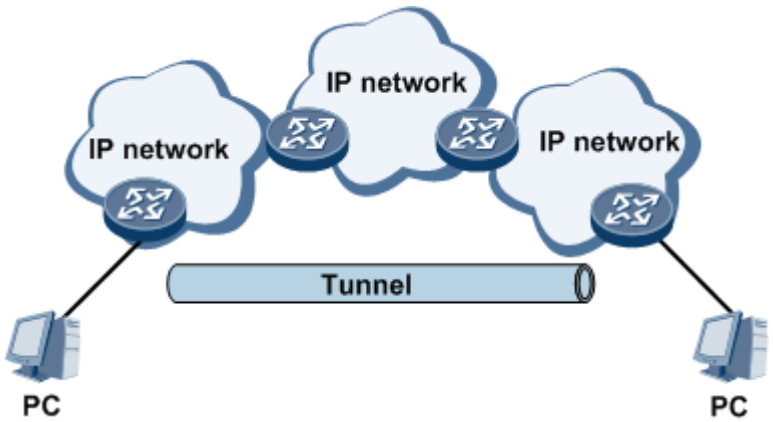
  The Key field contains a four-byte number, which is inserted into the GRE header during packet encapsulation. Packets of the same traffic flow have the same Key field. When decapsulating packets, a tunnel endpoint identifies packets of the same traffic flow based on the Key field.

  The authentication succeeds only if the Key fields set on both endpoints of the tunnel are the same. Otherwise, the packet is discarded. "Same" means that the Key field is not set on either endpoint, or is set to the same value on both endpoints.

9.6.4 GRE Applications

Workaround for Networks that Support a Limited Number of Hops

Figure 9-82 Enabling communication in a larger network scope



On the network shown in **Figure 9-82**, the IP protocol runs on the network. Assume that the hop limit of the IP protocol is 255. If the hop count between two PCs is greater than 255, the PCs cannot communicate. After a tunnel is used on the network, the intermediate hops between the two tunnel endpoints are hidden so that the data can traverse a larger network scope.

 Copyright © Huawei Technologies Co., Ltd.

Setting Up a VPN Across Physically Separate Networks

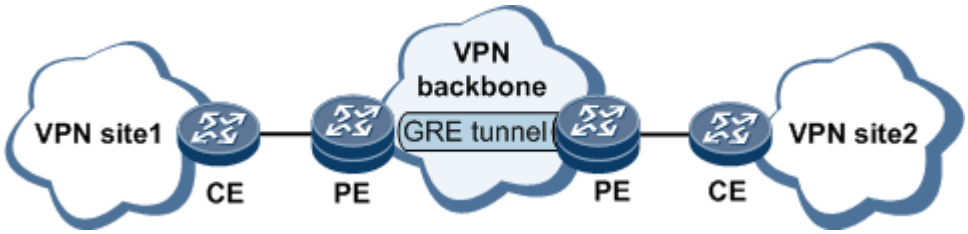
With GRE tunnels, you can set up a VPN across networks physically separated by a WAN.

Assume that Site 1 and Site 2 are deployed in different cities. By setting up a GRE tunnel between the PEs, you can connect the two separate networks to establish a VPN.

GRE, which applies to both L2VPNs and L3VPNs, can be used in either CPE-based VPNs or network-based VPNs:

In a network-based VPN scenario, both ends of the GRE tunnel reside on PEs, as shown in **Figure 9-83**.

Figure 9-83 GRE in a network-based VPN

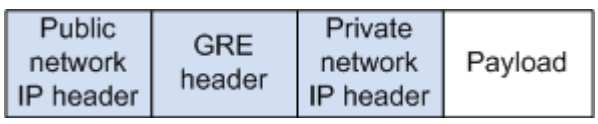


Usually, the VPN backbone network uses label switched paths (LSPs) as public network tunnels. If the core devices (P devices) on the backbone network do not support MPLS, LSPs cannot be used as public network tunnels. Then, you can use the GRE tunnel in place of the LSP to provide Layer 3 or Layer 2 VPN solutions at the core network. **Figure 9-84** shows the format of a private network packet transmitted on the VPN backbone network.

Figure 9-84 Format of a GRE packet that contains an MPLS label

GRE tunnels can also be used as non-MPLS VPN backbone tunnels. In this case, private network packets cannot contain MPLS labels when being transmitted on the VPN backbone network. **Figure 9-85** shows the format of such a packet.

Figure 9-85 Format of a GRE packet that does not contain an MPLS label



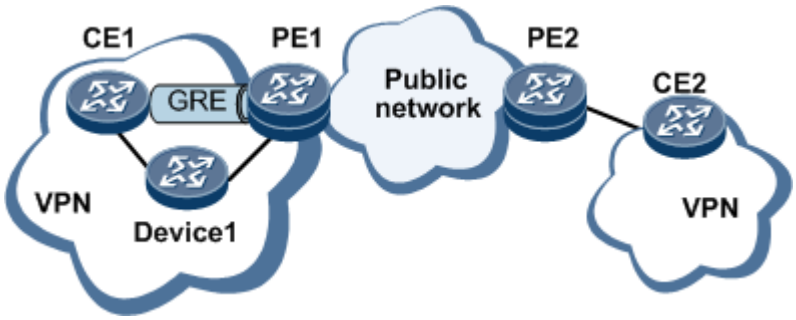
 Copyright © Huawei Technologies Co., Ltd.

Communication Across a WAN

Static routes can be used to connect a private network to a public network or another physically separate private network. This configuration applies to only IPv4 networks.

GRE on a LAN

Figure 9-86 GRE on a LAN

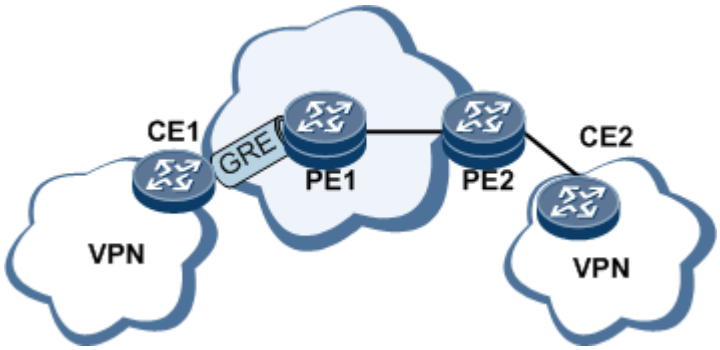


As shown in **Figure 9-86**, both the source and destination addresses of the GRE tunnel are private addresses for PE1.

GRE Across a WAN

In this networking, CEs and PEs must have public addresses. In addition, CE1 and PE1 must have a route to each other.

Figure 9-87 GRE across a WAN



The traffic destined from the LAN connected to CE1 for remote networks must pass through the GRE tunnel. Therefore, in the routing table on CE1, the outgoing interface must be set to the source interface of the GRE tunnel for all routes pointing to remote sites, and the next hop must be set to the remote tunnel interface. If multiple LANs are connected to CE1, you can select a LAN to bind it to the source tunnel interface.

 Copyright © Huawei Technologies Co., Ltd.

9.6.5 GRE Configuration

This section describes how to configure GRE.

9.6.5.1 Example for Configuring a Static Route for GRE

This section describes how to configure a static route for generic routing encapsulation (GRE) to transmit user traffic over the GRE tunnel. A static route is required between the device and its connected client.

Context

As shown in **Figure 9-88**, DeviceA (OLT), DeviceB (OLT), and DeviceC (router) belong to the backbone VPN and Open Shortest Path First (OSPF) runs between them.

A direct link is required between DeviceA and DeviceC. To meet such a requirement, configure a GRE tunnel between DeviceA and DeviceC, create a static route, and specify the tunnel interface as the outgoing interface of the static route so that PC1 and PC2 can communicate with each other.

PC1 and PC2 use DeviceA and DeviceC respectively as their default gateways.

Figure 9-88 Networking of configuring a static route for GRE

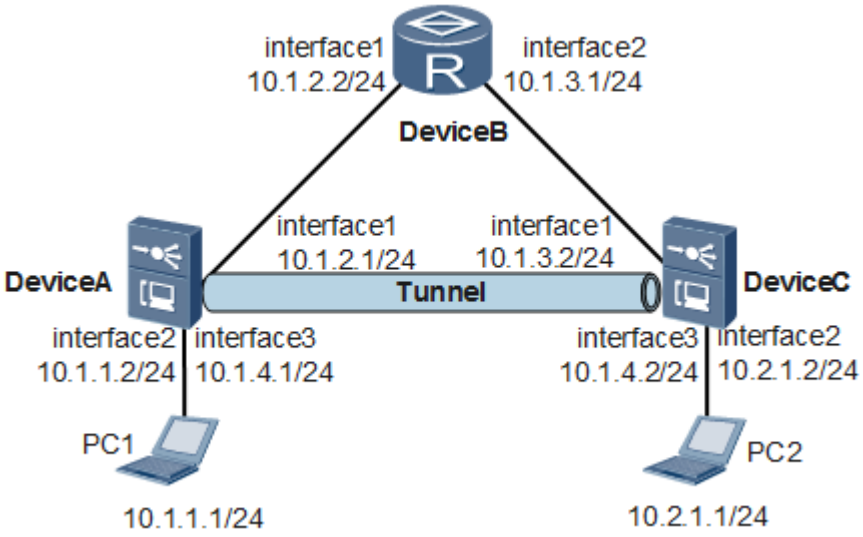


Table 9-15 Data planning

Configuration Item	Parameter	Value
Layer 3 interface	VLAN port	vlan 10 vlan 2

 Copyright © Huawei Technologies Co., Ltd.

Configuration Item	Parameter	Value
OSPF route	Device IP address	DeviceA: 10.1.2.1 DeviceB (interface 1): 10.1.2.2 DeviceB (interface 2): 10.1.3.1 DeviceC: 10.1.3.2
GRE tunnel	Tunnel IP address	10.1.4.1 10.1.4.2
	Address of the device at each end of the tunnel	DeviceA: 10.1.2.1 DeviceC: 10.1.3.2

Prerequisite

To complete the configuration, you need the following data:

- Data required for running OSPF by referring to **11.12.6.3 Configuration Example of OSPF**.
- Source and destination addresses and IP address of the tunnel interface at each end of the GRE tunnel.

Procedure

Step 1 Configure the interfaces and routes.

Step 2 Configure OSPF by referring to Configuration Example of OSPF to ensure connectivity between NEs.

Step 3 Configure a GRE tunnel on DeviceA and DeviceC.

Step 4 Configure GRE binding and paths.

----End

Example

1. 1. Configure the interfaces and routes.

   Configure the route between DeviceA and DeviceC because GRE tunnel transmission is based on Layer 3 IP forwarding.

   - Configure a VPN on DeviceA.

     ```
     huawei(config)#ip vpn-instance vpn1
     huawei(config-upn-instance-vpn1)#ipv4-family
     huawei(config-upn-instance-vpn1-af-ipv4)#quit
     huawei(config-upn-instance-vpn1)#quit
     ```

   - Configure a Layer 3 interface on DeviceA.

     ```
     huawei(config)#vlan 10 smart
     huawei(config)#port vlan 10 0/9 1
     huawei(config)#interface vlanif 10
     huawei(config-if-vlanif10)#ip address 10.1.2.1 24
     huawei(config-if-vlanif10)#quit
     ```

 Copyright © Huawei Technologies Co., Ltd.

- Configure OSPF on DeviceA.

```
huawei(config)#router id 10.10.10.10
huawei(config)#ospf
huawei(config-ospf-1)#area 0
huawei(config-ospf-1-area-0.0.0.0)#network 10.1.2.1 0.0.0.255
huawei(config-ospf-1-area-0.0.0.0)#quit
huawei(config-ospf-1)#quit
```

- Configure a VPN on DeviceC.

```
huawei(config)#ip vpn-instance vpn1
huawei(config-upn-instance-vpn1)#ipv4-family
huawei(config-upn-instance-vpn1-af-ipv4)#quit
huawei(config-upn-instance-vpn1)#quit
```

- Configure a Layer 3 interface on DeviceC.

```
huawei(config)#vlan 10 smart
huawei(config)#port vlan 10 0/9 1
huawei(config)#interface vlanif 10
huawei(config-if-vlanif10)#ip address 10.1.3.2 24
huawei(config-if-vlanif10)#quit
```

- Configure OSPF on DeviceC.

```
huawei(config)#router id 10.20.20.20
huawei(config)#ospf
huawei(config-ospf-1)#area 0
huawei(config-ospf-1-area-0.0.0.0)#network 10.1.3.2 0.0.0.255
huawei(config-ospf-1-area-0.0.0.0)#quit
huawei(config-ospf-1)#quit
```

- On DeviceB, set the address of the interface connected to DeviceA to 10.1.2.2 and that of the interface connected to DeviceC to 10.1.3.1.

  Then, configure routes on DeviceB. For details, see *Router Configuration*.

2. Configure OSPF by referring to **11.12.6.3 Configuration Example of OSPF** to ensure connectivity between NEs.
3. Configure a GRE tunnel on DeviceA and DeviceC.
   - Configure a GRE tunnel on DeviceA.

```
huawei(config)#interface tunnel 11
huawei(config-if-tunnel11)#ip binding vpn-instance vpn1
huawei(config-if-tunnel11)#tunnel-protocol gre
huawei(config-if-tunnel11)#ip address 10.1.4.1 255.255.255.0
huawei(config-if-tunnel11)#source 10.1.2.1
huawei(config-if-tunnel11)#destination 10.1.3.2
```

   - Configure a GRE tunnel on DeviceC.

```
huawei(config)#interface tunnel 11
huawei(config-if-tunnel11)#ip binding vpn-instance vpn1
huawei(config-if-tunnel11)#tunnel-protocol gre
huawei(config-if-tunnel11)#ip address 10.1.4.2 255.255.255.0
huawei(config-if-tunnel11)#source 10.1.3.2
huawei(config-if-tunnel11)#destination 10.1.2.1
```

4. Configure GRE binding and paths.
   - Create a VLAN and implement GRE binding on DeviceA.

```
huawei(config)#vlan 2 smart
huawei(config)#port vlan 2 0/9 0
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip binding vpn-instance vpn1
huawei(config-if-vlanif2)#ip address 10.1.1.2 24
huawei(config-if-vlanif2)#quit
```

   - Set the next hop to tunnel 11 on DeviceA. The destination network is 10.2.1.0.

```
huawei(config)#ip route-static 10.2.1.0 255.255.255.0 tunnel 11
```

   - Create a VLAN and implement GRE binding on DeviceC.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#vlan 2 smart
huawei(config)#port vlan 2 0/9 0
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip binding vpn-instance vpn1
huawei(config-if-vlanif2)#ip address 10.2.1.2 24
huawei(config-if-vlanif2)#quit
```

- Create a VPN route and set the next hop to tunnel 11 on DeviceC. The destination network is 10.1.1.0.

```
huawei(config)#ip route-static 10.1.1.0 255.255.255.0 tunnel 11
```

9.6.5.2 Example for Configuring a L3VPN over GRE Tunnel

L3VPN over GRE uses public GRE tunnels to carry L3VPN services. L3VPN/MPLS uses label switched paths (LSPs) as the public network tunnels. If the core router (P router) in the backbone network does not support MPLS, the LSPs cannot be used as the public network tunnels. In such a case, GRE tunnels can be used to implement L2VPN or L3VPN solutions in the core network.

Networking

As shown in **Figure 9-89**:

- PEs distribute private routes carrying labels through the Multiprotocol Extensions for BGP (MP-BGP) protocol.
- VPN packets can be transmitted to the peer PE after GRE encapsulation and two-layer inner labels.
- Tunnel policies are applied on the PE and GRE tunnels are specified for VPN packet forwarding.

Figure 9-89 Networking of configuring a L3VPN over GRE tunnel

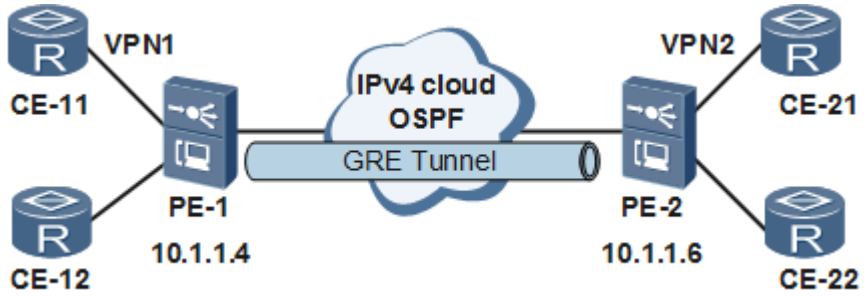


Table 9-16 Data planning

Configuration Item	Parameter	Value
Tunnel policy	Tunnel policy name	policy1
IP address	PE1 IP address	10.1.1.4
	PE2 IP address	10.1.1.6
GRE tunnel	GRE tunnel	tunnel11

 Copyright © Huawei Technologies Co., Ltd.

Prerequisite

To complete the configuration, you need the following data:

- GRE license that has been applied for and installed
- Source and destination addresses of the GRE tunnel
- Name of the VPN instance created on the PE

Procedure

Step 1 Configure GRE tunnels by referring to "Example for Configuring a Static Route for GRE".

Step 2 Create a VPN and tunnel policy, and bind the VPN instance to the tunnel policy.

Step 3 Set up the adjacency relationship between PE devices through the BGP protocol.

Step 4 Set up the tunnel by using the adjacency relationship. The source and destination addresses of the tunnel must be the same as those of the BGP peers.

Step 5 Create a Layer 3 interface and bind it to a VPN.

----End

Example

1. 1. Configure a VPN and tunnel policy.
   - Create a tunnel policy.
     ```
     huawei(config)#tunnel-policy policy1
     huawei(config-tunnel-policy-policy1)#tunnel select-seq gre load-balance-number 1
     huawei(config-tunnel-policy-policy1)#quit
     ```
   - Create a VPN and bind it to the tunnel policy.
     ```
     huawei(config)#ip vpn-instance vpn1
     huawei(config-vpn-instance-vpn1)#ipv4-family
     huawei(config-vpn-instance-vpn1-af-ipv4)#route-distinguisher 77:1
     huawei(config-vpn-instance-vpn1-af-ipv4)#tnl-policy policy1
     huawei(config-vpn-instance-vpn1-af-ipv4)#vpn-target 300:11 export-extcommunity
     huawei(config-vpn-instance-vpn1-af-ipv4)#vpn-target 400:22 import-extcommunity
     huawei(config-vpn-instance-vpn1-af-ipv4)#quit
     huawei(config-vpn-instance-vpn1)#quit
     ```
2. Set up the adjacency relationship between PE devices through the BGP protocol.
   - Configure the BGP protocol on the Layer 3 interface of PE1.
     ```
     huawei(config)#interface loopback 0
     huawei(config-if-loopback0)#ip address 10.1.1.4 255.255.255.255
     huawei(config-if-loopback0)#quit
     huawei(config)#bgp 2345
     huawei(config-bgp)#peer 10.1.1.6 as-number 2345
     huawei(config-bgp)#peer 10.1.1.6 connect-interface loopback 0
     huawei(config-bgp)#ipv4-family unicast
     huawei(config-bgp-af-ipv4)#peer 10.1.1.6 enable
     huawei(config-bgp-af-ipv4)#quit
     huawei(config-bgp)#ipv4-family vpnv4
     huawei(config-bgp-af-vpnv4)#policy vpn-target
     huawei(config-bgp-af-vpnv4)#peer 10.1.1.6 enable
     huawei(config-bgp-af-vpnv4)#quit
     huawei(config-bgp)#ipv4-family vpn-instance vpn1
     huawei(config-bgp-vpn1)#import-route direct
     huawei(config-bgp-vpn1)#quit
     huawei(config-bgp)#quit
     ```

 Copyright © Huawei Technologies Co., Ltd.

- Configure the BGP protocol on the Layer 3 interface of PE2.

```
huawei(config)#interface loopback 0
huawei(config-if-loopback0)#ip address 10.1.1.6 255.255.255.255
huawei(config-if-loopback0)#quit
huawei(config)#bgp 2345
huawei(config-bgp)#peer 10.1.1.4 as-number 2345
huawei(config-bgp)#peer 10.1.1.4 connect-interface loopback 0
huawei(config-bgp)#ipv4-family unicast
huawei(config-bgp-af-ipv4)#peer 10.1.1.4 enable
huawei(config-bgp-af-ipv4)#quit
huawei(config-bgp)#ipv4-family vpnv4
huawei(config-bgp-af-vpnv4)#policy vpn-target
huawei(config-bgp-af-vpnv4)#peer 10.1.1.4 enable
huawei(config-bgp-af-vpnv4)#quit
huawei(config-bgp)#ipv4-family vpn-instance vpn1
huawei(config-bgp-vpn1)#import-route direct
huawei(config-bgp-vpn1)#quit
huawei(config-bgp)#quit
```

3. Configure a GRE tunnel.
   - Configure a GRE tunnel on PE1.

```
huawei(config)#interface tunnel 11
huawei(config-if-tunnel11)#tunnel-protocol gre
huawei(config-if-tunnel11)#ip address 10.1.4.1 255.255.255.0
huawei(config-if-tunnel11)#source 10.1.1.4
huawei(config-if-tunnel11)#destination 10.1.1.6
huawei(config-if-tunnel11)#quit
```

   - Configure a GRE tunnel on PE2.

```
huawei(config)#interface tunnel 11
huawei(config-if-tunnel11)#tunnel-protocol gre
huawei(config-if-tunnel11)#ip address 10.1.4.2 255.255.255.0
huawei(config-if-tunnel11)#source 10.1.1.6
huawei(config-if-tunnel11)#destination 10.1.1.4
huawei(config-if-tunnel11)#quit
```

4. Configure OSPF by referring to **11.12.6.3 Configuration Example of OSPF**.
5. Bind the Layer 3 interface to the VPN.

```
huawei(config)#vlan 100 smart
huawei(config)#interface vlanif 100
huawei(config-if-vlanif100)#ip binding vpn-instance vpn1
huawei(config-if-vlanif100)#ip address 10.10.10.10 255.255.255.0
huawei(config-if-vlanif100)#quit
```

 Copyright © Huawei Technologies Co., Ltd.

10 Layer 2 Forwarding

10.1 MAC Address Management
MAC address management is a basic Layer 2 management feature.

10.2 VLAN
Virtual local area network (VLAN) is a technology used to form virtual workgroups by logically grouping the devices of a LAN.The VLAN management feature facilitates carriers' service planning.

10.3 Service Flow

10.4 Service Port Bundle

10.5 Layer 2 Forwarding Policy

10.6 Layer 2 User Bridging
Layer 2 user bridging implements Layer 2 data exchange among users under the same MA5800.

10.1 MAC Address Management

MAC address management is a basic Layer 2 management feature.

10.1.1 What Is MAC Address Management

Definition

MAC address management is a basic Layer 2 management feature that enables system administrators to use the functions listed in the following table.

 Copyright © Huawei Technologies Co., Ltd.

Table 10-1 Sub-functions of MAC address management

Sub-function of MAC Address Management	Description	Remarks
Setting the MAC address aging time	After a system administrator sets the MAC address aging time, the system periodically checks for aged dynamic MAC addresses. If the system detects no packets, whether sent or received, carrying specified source MAC addresses within one or two times of the aging time, the system deletes the MAC address from the MAC address table.	The system saves the MAC address table in its buffer. MAC address entries specify the mapping between the MAC addresses of device ports, port numbers, and VLAN IDs. When forwarding frames, the system consults the MAC address table according to the destination MAC addresses and VLAN IDs of these frames and quickly identifies the egress for them. This function prevents frame broadcast. MAC addresses in a MAC address table can be either manually configured or dynamically learned by the system.
Limiting the number of learnable dynamic MAC addresses	The system allows the system administrator to configure the number of learnable dynamic MAC addresses on a port or a service flow. When the number of learned MAC addresses reaches the maximum number, the port or service flow does not learn any new MAC addresses.	None
Setting the static MAC address	To connect a port to a device with a specified MAC address, the system administrator configures a static MAC address and VLAN on a port or configure a static MAC address on a service port. The system then forwards data according to this static MAC address.	None

Benefits

Benefits for Carriers

- The system ages dynamic MAC addresses to ensure timely updates of the MAC address table. If the MAC address table is full and not updated, the system will fail to learn new MAC addresses and will consequently fail to forward data.

 Copyright © Huawei Technologies Co., Ltd.

- By limiting the number of learnable dynamic MAC addresses, the system administrator can limit the number of MAC addresses that can be used to enter the network and hence alleviate the load of network devices.
- By configuring static MAC addresses, the system administrator prohibits unauthorized users from accessing the system.

  NOTE

  - If a malicious user sets source MAC addresses of attack packets to the MAC addresses of authorized users and accesses the system through another port on the device, the device learns an incorrect MAC address entry and forwards to the malicious user the packets that were originally destined to an authorized user. To avoid this situation, the system administrator can manually add specific MAC address entries to the MAC address table to bind user devices to ports. This operation protects user data from MAC spoofing and enhances port security. The manually added MAC address entries (static MAC address entries) have higher priorities than automatically generated MAC address entries.
  - The static MAC addresses, however, do not automatically update according to network device changes. The system administrator needs to modify the static MAC addresses manually. Using a large number of MAC addresses will require higher costs for network maintenance.

Benefits for Subscribers

Improved user security: After the system administrator sets the static MAC address of a service port and sets the maximum number of learnable MAC addresses to 0, the port receives only user data carrying the specified static MAC address.

10.1.2 MAC Address Management Process

MAC address management includes MAC address table establishment and management

Establishing MAC Address Tables

The system establishes a MAC address table by learning source MAC addresses or after users configure static MAC address entries.

- **MAC address learning**
  - When the OLT functions as a Layer 2 switching device, it learns MAC addresses in the distributed mode. Specifically, each board learns the source MAC address of packets sent from the board of its own and then forwards packets according to their destination MAC addresses. The learned MAC addresses are stored in the system buffer. Generally, the system buffer can hold a limited number of MAC address entries. If all these entries are filled in, no more MAC addresses can be learned.
  - Configuration command: **mac-address learning**

  NOTE

  The MAC address learning function takes effect only for unicast MAC addresses, but not for multicast or broadcast MAC addresses.

- **Configuration of static MAC address entries**
  - A user can manually configure static MAC address entries in which user device MAC addresses are bound to ports. After this configuration, the packets whose MAC addresses are included in the MAC address entries

 Copyright © Huawei Technologies Co., Ltd.

are always forwarded through the bound ports. This configuration improves the efficiency for forwarding packets and improves the security of ports because it denies access from unauthenticated users. This method of establishing MAC address tables is widely used in private networks.

NOTE

The same static MAC address can be configured on an upstream port belonging to different VLANs.

– Configuration command: **mac-address static**

The following table shows an example of a simplified MAC address table established by configuring static MAC address entries. The table lists the mapping between MAC addresses, ports, and VLAN IDs.

Table 10-2 Simplified MAC address table

VLAN ID	MAC Address	Port Number	MAC Address Attribute	Forwarding Attribute
2	00e0-fc00-0011	0/1/1	Dynamic	Forward
3	00e0-fc00-0033	0/1/2	Static	Forbid

Managing MAC Address Tables

When managing MAC address tables, users can configure MAC-related attributes as allowed by system resources and network security policies against potential risks. The optimized MAC address tables can better meet requirements of a live network. These MAC-related attributes are as follows:

- **Maximum number of MAC addresses learned based on service flows**
  – After the number of access users reaches the limit, no new access user addresses will be learned. This attribute setting applies to networks, such as residential access networks and low-security internal enterprise networks, that have fixed access users but are not sufficiently secure.

    NOTE

    - The maximum number of MAC addresses is independent of the number of manually added static MAC addresses.
    - After a static MAC address is configured for a user port and the maximum number of learnable dynamic MAC addresses is set to 0, the port receives only packets that contain the configured static MAC address. In this case, the MAC address is bound to the port. This attribute setting helps improve user data security.

  – Specifically, after a service flow is added to a bundle, the system uses the maximum number for the bundle as that for the service flow. The maximum number of learnable MAC addresses for the entire bundle is the same as that for each service flow in the bundle. When the maximum number for any service flow in a bundle changes, the maximum number for each bundle and that for each service flow in the bundle change as

 Copyright © Huawei Technologies Co., Ltd.

well. The sum of MAC addresses learned for each service flow in a bundle cannot exceed the maximum number of learnable MAC addresses for the bundle.

– Configuration command: **mac-address max-mac-count**

- **Setting the function of sensing excess MAC addresses**

  When a lot of MAC addresses are learnt by the system, it is difficult for trouble locating. When the function of sensing excess MAC addresses is enabled, the system queries MAC addresses every 15 minutes and determines whether an alarm needs to be reported according to the query result. If the query result exceeds the upper threshold for sensing excess MAC addresses set by users, an excess MAC address alarm is generated. If the query result is smaller than the lower threshold for sensing excess MAC addresses set by users, a fault clearing alarm is generated.

  Configuration command: **overload-aware mac-address**

- **MAC address aging**

  Generally, the system automatically establishes a MAC address table by learning source MAC addresses. The system can age out dynamic MAC addresses. When the system does not receive packets from or send packets to a device within a specified period, the system deletes the MAC address entry of the device from the MAC address table. The established MAC address table has to be updated according to network changes. However, after the network topology changes, the dynamic MAC address entries will not be automatically updated in a timely manner. Then the system cannot learn more MAC addresses and user data cannot be forwarded as expected. MAC address aging is intended to resolve the issue.

  NOTE

  Disable MAC address aging if no MAC address aging issue has occurred.

  The MAC address learning function can be configured for the control board and service boards separately.

Item	Description	Configuration Command
Specify the aging time of a MAC address	After a system administrator sets the MAC address aging time, the system periodically checks for aged dynamic MAC addresses. If the system detects no packets, whether sent or received, carrying specified source MAC addresses within one or two times of the aging time, the system deletes the MAC address from the MAC address table. Then the system creates another MAC address table by learning MAC addresses.	**mac-address timer**

 Copyright © Huawei Technologies Co., Ltd.

Item	Description	Configuration Command
Sets the aging mode for a MAC address	After you set the MAC address aging mode, the control board and service board age MAC addresses accordingly. • Bidirectional aging mode: The system ages MAC address entries according to the source MAC address learning and receiving of packets destined for this MAC address. Specifically, if the system neither learns a source MAC address nor receives packets destined for this MAC address within an aging period, the system ages the MAC address entry. Bidirectional aging helps reduce the CPU usage required for aging and learning MAC addresses. • Unidirectional aging mode: The system ages MAC address entries according to the source MAC address learning. Specifically, if the system does not learn a MAC address again within an aging period, the system ages the MAC address entry.	**mac-address aging-mode**

NOTE

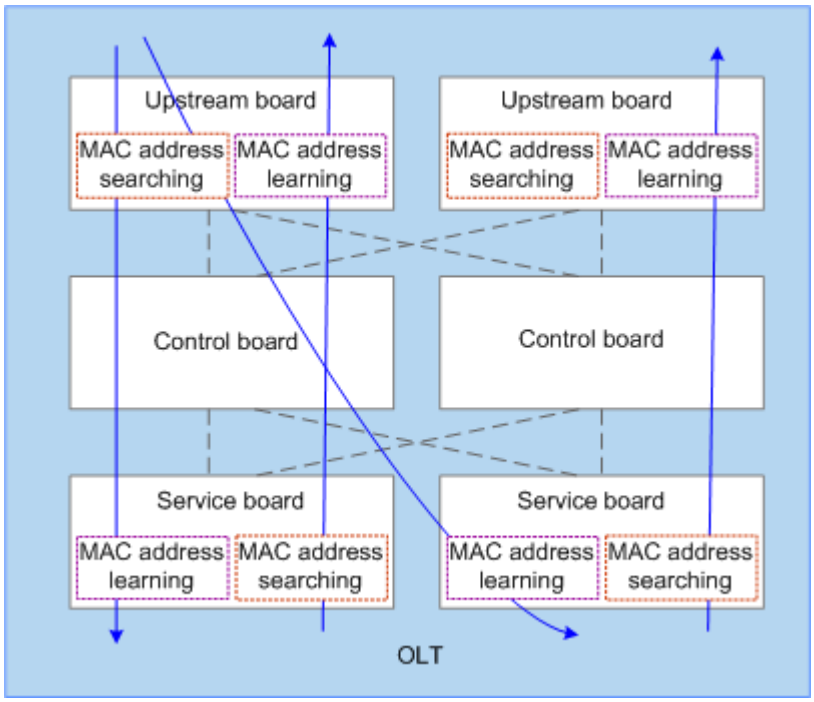
- MAC address hash conflict: HASH algorithm is an algorithm commonly used by the forwarding chip in order to accelerate finding forwarding entries. Based on VLAN + MAC forwarding mode, the system make the hash operation on VLAN + MAC address by a HASH algorithm when learning the MAC address to get a hash leaky bucket index. If the number of MAC address table entries of the leaky bucket corresponding to the index has reached the maximum, this MAC address that is operated would not be learned by the system. It is called to generate MAC address hash conflict. (The number of MAC address table entries in each leaky bucket is fixed. It may be different for different control board).
- The system uses the hash algorithm to calculate MAC address tables, so a hash conflict issue may occur.

10.1.3 Smooth MAC Address Expansion

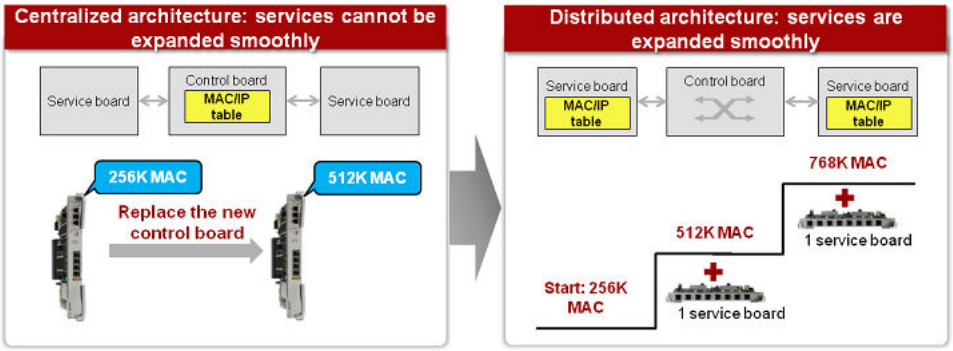
The MA5800 is designed in distributed architecture. Compared with devices in centralized architecture, MAC addresses of the MA5800 can be smoothly expanded.

The MA5800 learns and forwards MAC addresses as an entire system. Specifically, its egress board learns MAC addresses and its ingress board finds and forwards MAC addresses.

 Copyright © Huawei Technologies Co., Ltd.



Compared with devices in centralized architecture, the MA5800 has the following advantages:



- Centralized architecture:
  - MAC address learning and forwarding are performed on the control board. For any MAC address expansion, the control board needs to be replaced.
  - In the initial stage of network construction, usage of the control board having high specifications is costly.
- Distributed architecture:
  - MAC address expansion is implemented by service board expansion but the control board does not need to be replaced.
  - MAC addresses are expanded smoothly, meeting customer requirements in different stages.

 Copyright © Huawei Technologies Co., Ltd.

10.2 VLAN

Virtual local area network (VLAN) is a technology used to form virtual workgroups by logically grouping the devices of a LAN.The VLAN management feature facilitates carriers' service planning.

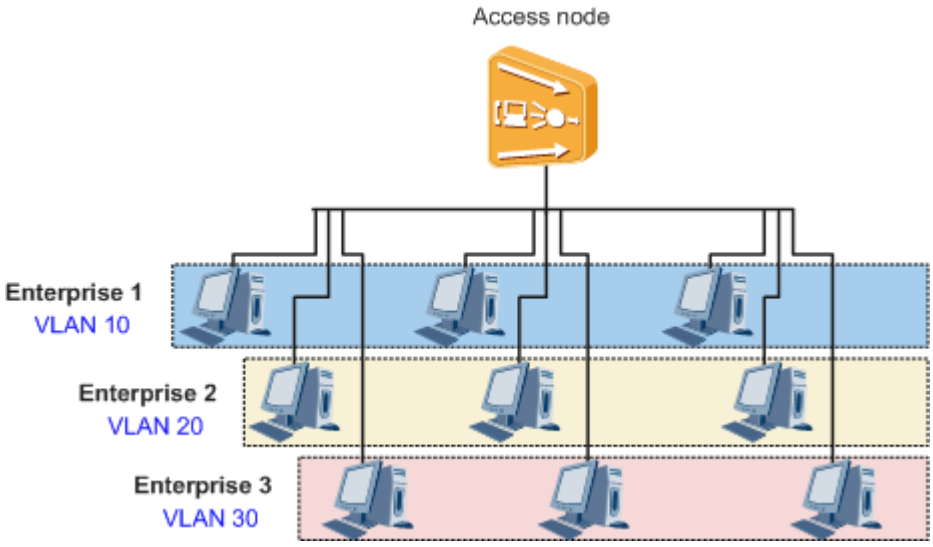
10.2.1 What Is VLAN

VLAN is a communications technology that divides a physical LAN into multiple logical broadcast domains (multiple VLANs). NEs in a VLAN can communicate with each other but NEs in different VLANs cannot.

VLAN allows packets to be broadcast only within a single VLAN, preventing bandwidth waste caused by broadcast storms. Furthermore, VLAN enhances network security, because NEs in different VLANs cannot communicate with each other directly.

For example, for users of different enterprises in the same office building, it is too costly to build separate LANs for each enterprise, and insecure if the enterprises are to share the existing LAN of the building. VLAN resolves this dilemma.

Figure 10-1 VLAN application



Make different enterprise users belonging to different VLANs. In this way, the enterprise users share the LAN facilities and at the same time each have their own secure networks.

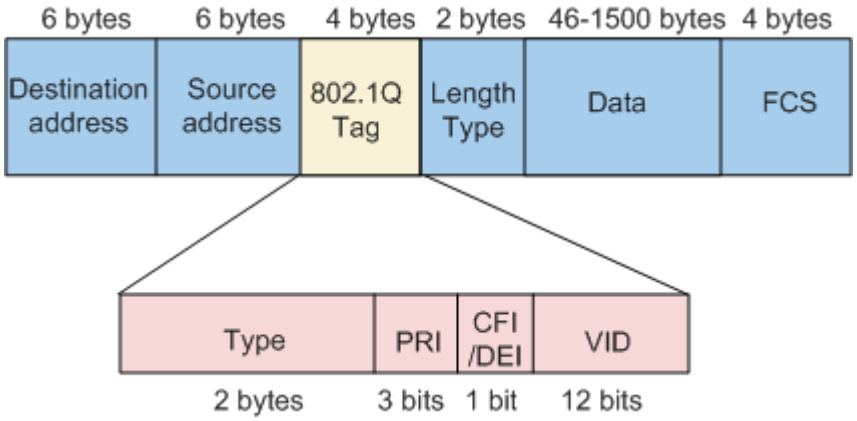
10.2.2 VLAN Basic Concepts

802.1Q Frame Format for a VLAN

The format of Ethernet frames is modified in the IEEE 802.1Q standard. Specifically, a 4-byte 802.1Q tag is inserted between the source MAC address field and the protocol type field, as shown in the following figure.

 Copyright © Huawei Technologies Co., Ltd.

Figure 10-2 802.1Q frame format for a VLAN



The 802.1Q tag consists of four fields:

- Type

  The Type field is 2 bytes long and indicates the frame type. The value 0x8100 of this field indicates that the frame carries an 802.1Q tag. Such a frame will be discarded if the receiving device does not support 802.1Q.

- PRI

  The PRI field is 3 bits long and indicates the frame priority. It ranges from 0 to 7. The larger the value, the higher the priority. In case of network congestion, the system preferentially sends data frames with a higher priority.

- CFI/DEI

  It is 1 bit long.

  – The CFI field indicates whether the MAC address is in canonical format. CFI 0 indicates that the MAC address is in the canonical format and CFI 1 indicates a non-canonical format. It is used to differentiate Ethernet frames, fiber distributed digital interface (FDDI) frames, and token ring frames. In an Ethernet network, CFI is 0.

  – The DEI field consists of one bit and represents the drop eligible indicator defined in the 802.1ad protocol. It is used to color the packet.

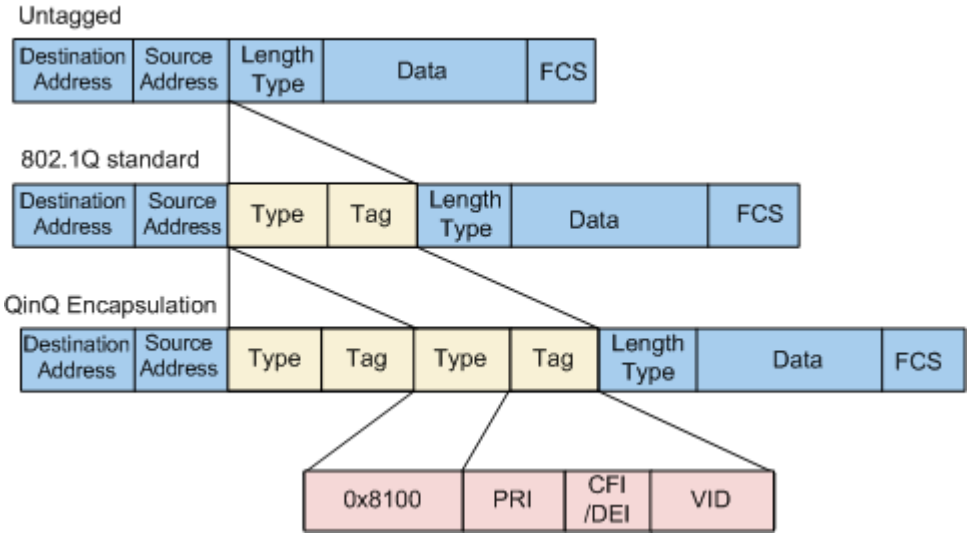
- VID

  VID is the shortened form of VLAN ID. It is 12 bits long and indicates the VLAN to which a frame belongs.

PRI, CFI, and VID together are called a VLAN tag, which is the basis for VLAN communication. Frames that do not contain an 802.1Q tag are called untagged frames. 802.1Q frames in QinQ encapsulation have two 802.1Q tags. The following figure shows the structures of untagged frames, standard 802.1Q frames, and 802.1Q frames in QinQ encapsulation.

 Copyright © Huawei Technologies Co., Ltd.

Figure 10-3 Structure comparison between untagged frame, standard 802.1Q frame, and 802.1Q frame in QinQ encapsulation



VLAN Types

Table 10-3 VLAN types

VLAN Type	The same points	Difference
Standard VLAN	• Ethernet ports in different VLANs are isolated from each other. • Ethernet ports can be communicated with each other. • A VLAN contains multiple upstream ports.	A standard VLAN does not contain a service port.
Smart VLAN		• A smart VLAN contains multiple service ports. • Service ports in a smart VLAN are isolated from each other. A VLAN provides access for multiple users and therefore saves VLAN resources.
MUX VLAN (Multiplex vlan)		• A MUX VLAN contains only one service port. • A service port in a MUX VLAN is isolated from a service port in another MUX VLAN.
Super VLAN	The super VLAN is based on Layer 3. A super VLAN can contain multiple sub VLANs. The type of the sub VLAN can be smart VLAN or MUX VLAN. The sub VLANs can communicate with each other by ARP proxy. The sub VLAN forwards the service at Layer 3 centrally by the Layer 3 interface of the super VLAN. A super VLAN aggregates Smart VLAN or MUX VLANs. For details, see **10.2.4 VLAN Aggregation (Super VLAN)**.	

Copyright © Huawei Technologies Co., Ltd.

The Smart VLAN restricts visits between users by isolating the service flows or service ports in the same VLAN.

The MUX VLAN realizes user isolation by dividing the user service flows or service ports into different VLANs.

VLAN Attributes

Table 10-4 VLAN Attributes

Attribute	Description	Services Supported by Service Flows Inside the VLAN	Transparent Transmission Capability
Common	A VLAN with the common attribute supports only one-tagged VLANs. This VLAN is used as a common Layer 2 VLAN. If it is used for Layer 3 forwarding, it needs a VLAN interface.	**Supports** DHCP Option82, DHCP Layer 2 relay, DHCP Layer 3 relay, PITP (PPPoE+), MAC anti-spoofing, IP anti-spoofing, IPTV, and ARP proxy.	Transparent transmission by hardware is general. The service board does not support transparent transmission by hardware for service flows having common and stacking VLANs and it captures DHCP, PPP, and IGMP packets globally (if the related features are enabled on a VLAN).
Stacking	A VLAN with the stacking attribute supports only two-tagged VLANs. Both the outer VLAN tag and inner VLAN tag of a stacking VLAN are allocated by the MA5800, which are used for dual VLAN authentication for the upper-layer BRAS and provide the wholesale service for the ISP.		
QinQ	A VLAN with the QinQ attribute supports one-tagged or multi-tagged VLANs, which is mainly used in transparent transmission and private line scenarios. If a multi-tagged VLAN is used, both the outer VLAN tag and inner VLAN tag are allocated on the MA5800. If a three-tagged VLAN is used, user-side packets carry one-tagged VLAN and the MA5800 adds the two-tagged VLAN.	**Does not support** DHCP Option82, DHCP Layer 2 relay, DHCP Layer 3 relay, PITP (PPPoE+), MAC anti-spoofing, IP anti-spoofing, IPTV, or ARP proxy.	Transparent transmission by hardware is powerful cause the service board supports transparent transmission by hardware for service flows having QinQ VLANs.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

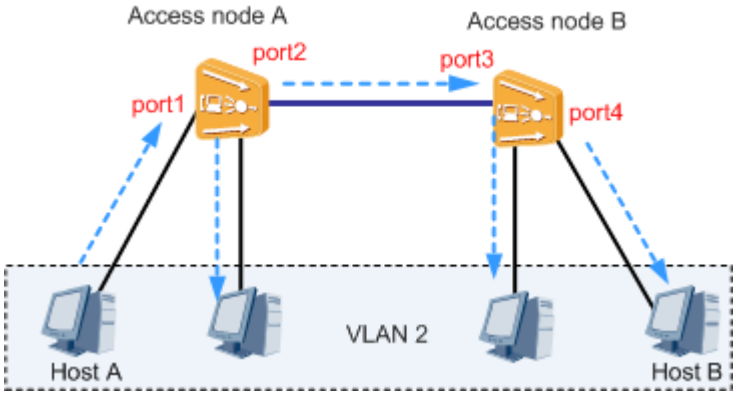
For details about QinQ VLANs and stacking VLANs, see **10.2.5 QinQ VLAN and Stacking VLAN**.

10.2.3 VLAN Communication Principle

Intra-VLAN Communication Principle

Users in a VLAN can communicate with each other. An access node implements the communication by distinguishing between the VLAN tags carried in Ethernet frames. In the following figure, there are 4 hosts in VLAN 2, and uses the example of host A and host B to illustrate communication inside a VLAN.

Figure 10-4 Intra-VLAN communication



1. Host A sends Ethernet frames to port 1 of access node A.
2. Port 1 of access node A attaches a VLAN tag (VID is filled in with VLAN 2) to the Ethernet frames.
3. Access node A sends the tagged Ethernet frames to ports (except port 1) that belong to VLAN 2.
4. Port 2 of access node A sends the tagged Ethernet frames to port 3 of access node B.
5. Access node B identifies the VLAN tag (VLAN 2) carried in the Ethernet frames and sends these frames to all ports of access node B that belong to VLAN 2.
6. Port 4 of access node B sends these Ethernet frames to host B.

Inter-VLAN Communication Principle

Hosts in different VLANs cannot communicate with each other at Layer 2. If these hosts need to communicate with each other, they need IP routes. The MA5800, serving as a Layer 3 switch, supports communication between hosts in different VLANs.

To ensure that the Layer 3 switch can transmit the data flow for the first time, the routing table must have correct routing entries. Therefore, configuring a Layer 3

 Copyright © Huawei Technologies Co., Ltd.

interface and a routing protocol on the Layer 3 switch is required. A VLANIF interface is a Layer 3 logical interface

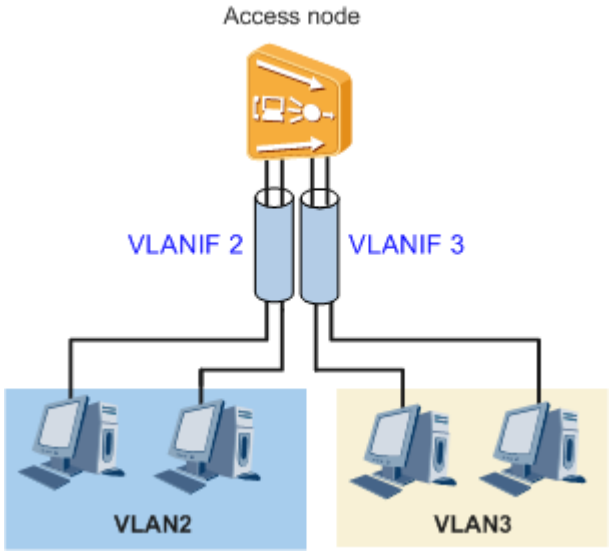
and its IP address can be manually specified.

NOTE

Layer 3 switching combines routing and switching technologies to implement routing inside a switch, thereby improving the performance of the entire network. After using the routing table to send the first data flow, a Layer 3 switch generates ARP table for MAC addresses and IP addresses. When the ARP entry exists for the corresponding destination IP, the system directly finds the out port by the ARP table entries and translates the destination MAC address, source MAC address, and VLAN.

In the following figure, the access node supports both VLAN 2 and VLAN 3. In this scenario, VLAN interface 2 and VLAN interface 3 can be created on the switch. After IP addresses and correct routes are configured for these two VLAN interfaces, VLAN 2 can communicate with VLAN 3.

Figure 10-5 Inter-VLAN communication



Default VLAN of an Upstream Port

The default VLAN (port default VLAN ID, or PVID) of a port is also called a native VLAN. If a default VLAN is configured on an Ethernet port, the processing rule of Ethernet port for different VLAN tag in sent packets is shown as follows.

Table 10-5 The processing rule of Ethernet port for different VLAN tag in received packets

VLAN Tag in Sent Packets on An Ethernet Port	Processing Rule of Ethernet port
Untagged	The Ethernet port tags the packets with the default VLAN before forwarding them.

 Copyright © Huawei Technologies Co., Ltd.

VLAN Tag in Sent Packets on An Ethernet Port	Processing Rule of Ethernet port
VLAN ID is the same as the default VLAN ID	The Ethernet port removes the VLAN tag from the packets before forwarding them.
VLAN ID is different from the default VLAN ID	The Ethernet port directly forwards these packets with the VLAN tag unchanged.

NOTE

If the MA5800 connects to an upper-layer router, packets sent by the router may or may not carry a VLAN tag, depending on whether the router can identify a VLAN tag. The default VLAN can be configured on the upstream port of the MA5800 to adapt the router.

10.2.4 VLAN Aggregation (Super VLAN)

Context

VLAN is widely applied to switching networks because of its flexible control of broadcast domains and convenient deployment. On a Layer-3 switch, interconnection between broadcast domains is implemented using one VLAN to correspond to a single Layer-3 logical interface. However, such an implementation may lead to inefficient use of IP addresses. VLAN aggregation, also known as a super VLAN, can solve this problem.

NOTE

Here takes an example of host address assignment in VLANs to show the inefficient use of IP addresses.

Example

The following figure shows a typical VLAN planning.

 Copyright © Huawei Technologies Co., Ltd.

Figure 10-6 Typical VLAN planning

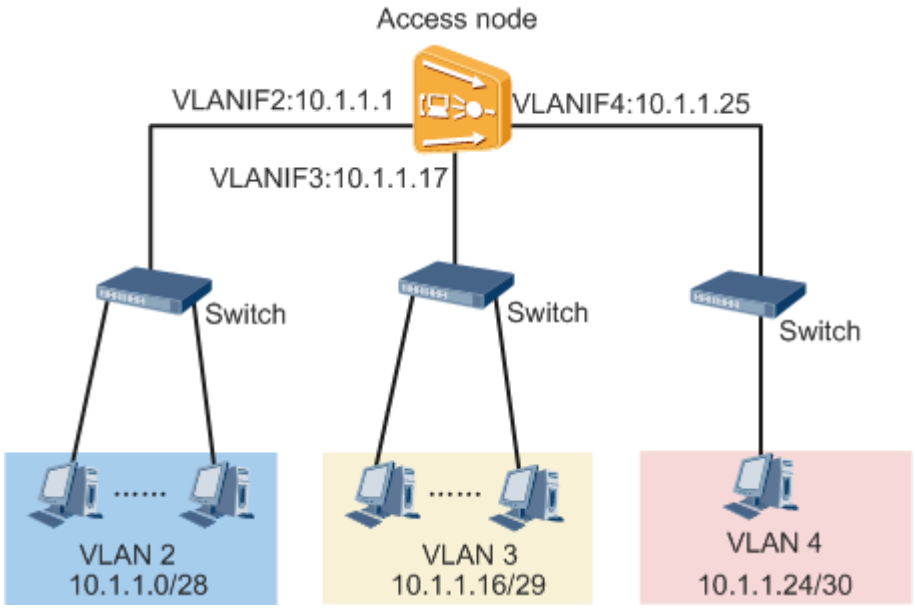


Table 10-6 Example of host address assignment in VLANs

VLAN	Subnet	Gateway Address	Number of Available Addresses	Number of Available Hosts	Actual Requirements
2	10.1.1.0/28	10.1.1.1	14	13	10
3	10.1.1.16/29	10.1.1.17	6	5	5
4	10.1.1.24/30	10.1.1.25	2	1	1

As shown in the preceding table,

- VLAN 2 requires 10 host addresses. Subnet 10.1.1.0/28 with a mask of 28 bits is assigned to VLAN 2. Here, 10.1.1.0 is the subnet address, and 10.1.1.15 is the directed broadcast address. These two addresses cannot be used as the host address. In addition, as the default gateway address of the subnet, 10.1.1.1 cannot be used as the host address, either. The other 13 addresses ranging from 10.1.1.2 to 10.1.1.14 can be used by the hosts. In this way, although VLAN 2 needs only 10 addresses, 13 addresses need to be assigned to it according to the subnetting principle.
- VLAN 3 requires five host addresses, and subnet 10.1.1.16/29 with a mask of 29 bits needs to be assigned to VLAN 3. VLAN 4 requires only one address, and subnet 10.1.1.24/30 with a mask of 30 bits needs to be assigned to VLAN 4.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

Directed broadcast address: In a network divided into different subnets, data packets are sent to all hosts in a subnet after a directed broadcast address is specified as the destination address of the packets. Subnet directed broadcasting is usually used to obtain data of a host in a subnet, such as the neighboring relationship.

In the above example, only 16 (10+5+1) addresses are actually required for all the VLANs. The VLAN planning, however, needs 28 (16+8+4) addresses according to the common VLAN addressing principle, even if the optimal addressing scheme is used. Therefore, nearly half of the addresses will be wasted. In addition, if VLAN 2 is accessed by only three hosts instead of ten, the remaining addresses will also be wasted because such addresses cannot be used by other VLANs.

This addressing plan is inconvenient for future network upgrades and expansion. If VLAN 4 needs addition two hosts and does not want to change the assigned IP addresses, and the addresses after 10.1.1.24 has been assigned to others, a new sub network with the mask length as 29 bits and a new VLAN need to be assigned for the new users of VLAN 4. As a result, VLAN 4's customers only have three hosts, but the customers are assigned to two different subnets in separate VLANs, which makes network management difficult.

In the above example, several IP addresses are used as subnet addresses, subnet directed broadcast addresses, and default addresses of subnet gateways. Such IP addresses cannot be used as host addresses in the VLAN. This address assigning mechanism greatly reduces addressing flexibility and causes address usage waste. VLAN aggregation is developed to resolve these issues.

Principle

VLAN aggregation, divides a physical network into broadcast domains by using VLANs so that different VLANs can belong to the same subnet. VLAN aggregation consists of two basic concepts, super VLAN and sub VLAN.

- Super VLAN: Super VLANs differ from common VLANs. In super VLANs, only Layer 3 interfaces are created and physical ports are not involved. A super VLAN can be regarded as a logical Layer 3 collection of many sub VLANs.
- Sub VLAN: Sub VLANs are used to isolate broadcast domains. In sub VLANs, only physical ports are contained and Layer 3 VLAN interfaces cannot be created. The Layer 3 switching between a sub VLAN and the external network is implemented through the Layer 3 interface of the super VLAN.

A super VLAN can contain one or more sub VLANs, each sub VLAN with different broadcast domains. The sub VLAN does not occupy an independent subnet segment. In the same super VLAN, IP addresses of hosts belong to the super VLAN's subnet segment, regardless of the mapping between hosts and sub VLANs.

The same Layer 3 interface is shared by sub VLANs, allowing fewer subnet IDs, default gateway addresses of the subnet, and directed broadcast addresses of the subnet to be used. In the mean time, different broadcast domains can use the unused addresses in the same subnet segment. As such, addressing becomes flexible and efficient.

NOTE

The following still uses **Table 10-6** as an example to explain the implementation principle of VLAN aggregation.

 Copyright © Huawei Technologies Co., Ltd.

Example

Supposing that user demands are the same. In VLAN 2, ten host addresses are demanded; in VLAN 3, five host addresses are demanded; in VLAN 4, one host address is demanded.

Create VLAN 10 and configure VLAN 10 as a super VLAN. Then assign subnet address 10.1.1.0/24 with a mask of 24 bits to VLAN 10, where 10.1.1.0 is the subnet ID and 10.1.1.1 is the gateway address of the subnet, as shown in the following figure. The corresponding sub VLAN address assignment of VLAN 2, VLAN 3, and VLAN 4 is shown in **Table 10-7**.

Figure 10-7 VLAN aggregation schematic diagram

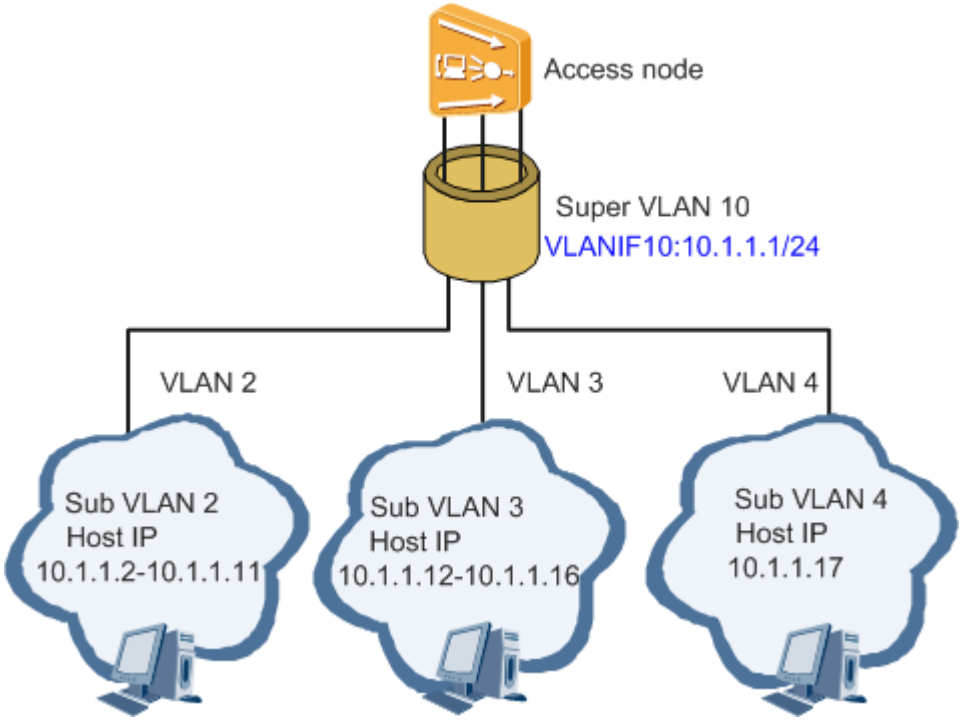


Table 10-7 Example of host address assignment in VLAN aggregation mode

VLAN	Subnet	Gateway Address	Number of Available Addresses	Available Addresses	Actual Requirements
2	10.1.1.0/24	10.1.1.1	10	10.1.1.2-10.1.1.11	10
3			5	10.1.1.12-10.1.1.16	5
4			1	10.1.1.17	1

In VLAN aggregation, sub VLANs are not divided according to the previous subnet border. Instead, their addresses are flexibly assigned among the super VLAN's

Copyright © Huawei Technologies Co., Ltd.

subnets according to the required number of hosts. VLAN 2, VLAN 3, and VLAN 4 share a subnet (10.1.1.0/24), a default gateway address of the subnet (10.1.1.1), and a directed broadcast address of the subnet (10.1.1.255). In this manner, the subnet ID (10.1.1.16, 10.1.1.24), the default gateway of the subnet (10.1.1.17, 10.1.1.25), and the directed broadcast address of the subnet (10.1.1.15, 10.1.1.23, and 10.1.1.27) can be used as host IP addresses.

In total, 16 addresses (10+5+1) are required for the three VLANs. In practice, in this subnet, a total of 16 addresses (10.1.1.2 to 10.1.1.17) are assigned to the three VLANs. A total of 19 IP addresses are used, that is, the 16 host addresses together with the subnet ID (10.1.1.0), the default gateway of the subnet (10.1.1.1), and the directed broadcast address of the subnet (10.1.1.255). In the network segment, 236 addresses (255 - 19) are available, which can be used by any host in the sub VLAN.

Communication Between Sub VLANs

VLAN aggregation ensures that different VLANs can use the IP addresses in the same subnet segment; however, this leads to an issue with Layer 3 forwarding between sub VLANs.

In common VLAN mode, the hosts of different VLANs can communicate with each other based on Layer 3 forwarding through their respective gateways. In VLAN aggregation mode, however, hosts in a super VLAN use IP addresses in the same network segment and share the same gateway address. Since hosts in different sub VLANs belong to the same subnet, they communicate with each other based on Layer 2 forwarding, not Layer 3 forwarding through a gateway. As a result, hosts in different sub VLANs cannot communicate with each other because the hosts are separated at Layer 2.

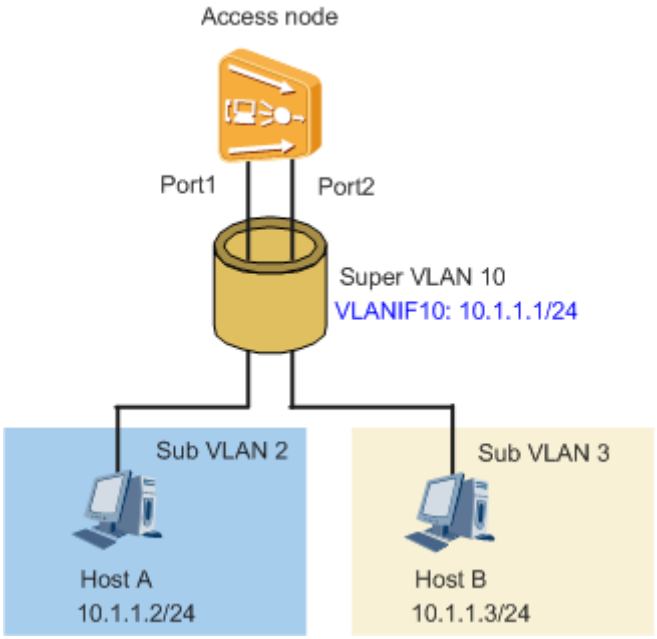
The Address Resolution Protocol (ARP) proxy resolves this issue.

- **Layer 3 communication between different sub VLANs**

  As shown in the following figure, the super VLAN, namely, VLAN 10, contains two sub VLANs, VLAN 2 and VLAN 3.

 Copyright © Huawei Technologies Co., Ltd.

Figure 10-8 Networking diagram of Layer 3 communication between different sub VLANs based on ARP proxy



Communication between host A in sub VLAN 2 and host B in sub VLAN 3 is implemented as follows:

NOTE

Suppose that host A's ARP table has no corresponding entry for host B, and the gateway between the sub VLANs is enabled with ARP proxy.

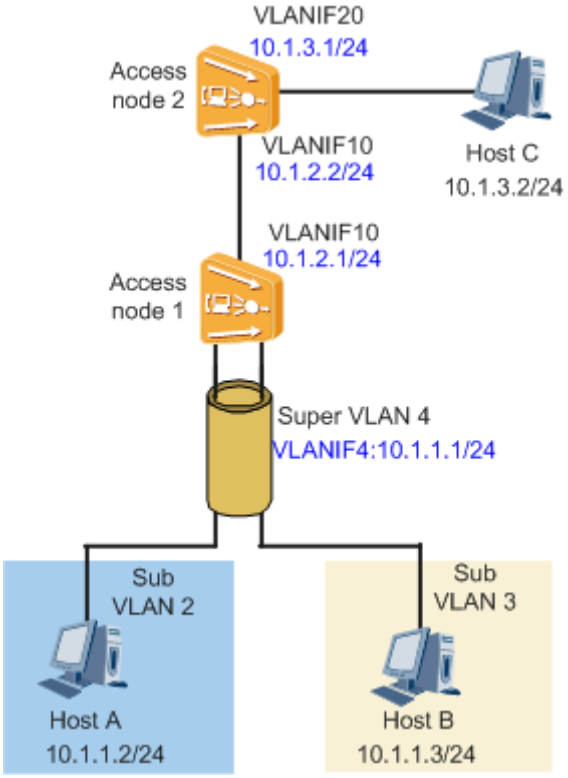
a. After comparing the IP address (10.1.1.3) of host B with its own IP address, host A learns that both IP addresses are in the same network segment (10.1.1.0/24), and its ARP table has no entry corresponding to host B.
b. Host A initiates ARP broadcasting to request host B's MAC address.
c. Host B is not in the broadcast domain of sub VLAN 2, and cannot receive the ARP request.
d. Since the gateway's ARP proxy is enabled between the sub VLANs, after receiving host A's ARP request, the gateway discovers that the IP address of host B (10.1.1.3) is the IP address of a directly-connected interface. The gateway then initiates ARP broadcasting to all other sub VLAN interfaces to request host B's MAC address.
e. After receiving an ARP request, host B replies with an ARP response.
f. After receiving host B's ARP response, the gateway replies to host A with the gateway's MAC address. Host A will regard this MAC address as host B's MAC address.
g. The ARP tables in both the gateway and host A have entries corresponding to host B.
h. To send packets to host B, host A initially sends packets to the gateway, and then the gateway performs Layer 3 forwarding to implement communication between host A and host B.

The process used by host B to send packets to host A works in the same way.

 Copyright © Huawei Technologies Co., Ltd.

- **Layer 3 communication between a sub VLAN and an external network**

  Figure 10-9 Networking diagram of Layer 3 communication between a sub VLAN and an external network



  As shown in the preceding figure, access node 1 is configured with super VLAN 4, sub VLAN 2, sub VLAN 3, and common VLAN 10. Access node 2 is configured with two common VLANs, VLAN 10 and VLAN 20. Host A in sub VLAN 2 that belongs to the super VLAN 4 needs to access host C in access node 2.

  NOTE

  Suppose that access node 1 is configured with a route to network segment 10.1.3.0/24, and access node 2 is configured with a route to network segment 10.1.1.0/24.

  a. After comparing the IP address (10.1.3.2) of host C with its IP address, host A learns that two IP addresses are not in the same network segment 10.1.1.0/24.
  b. Host A initiates ARP broadcasting to its gateway, requesting the gateway's MAC address.
  c. After receiving the ARP request, access node 1 identifies the correlation between the sub VLAN and the super VLAN, and replies with an ARP response to host A through sub VLAN 2. The source MAC address in the ARP response packet is the MAC address of VLANIF4 of super VLAN 4.
  d. Host A learns the gateway's MAC address.
  e. Host A sends the packet to the gateway, the destination MAC address being the MAC address of VLANIF4 of super VLAN 4, and the destination IP address being 10.1.3.2.

 Copyright © Huawei Technologies Co., Ltd.

f. After receiving the packet, access node 1 performs Layer 3 forwarding and sends the packet to access node 2, with the next hop address 10.1.2.2, and the outgoing interface VLANIF10.

g. After receiving the packet, access node 2 performs Layer 3 forwarding and sends the packet to host C through the directly-connected interface VLANIF20.

h. The response packet from host C reaches access node 1 after access node 2 performs Layer 3 forwarding.

i. After receiving the packet, access node 1 performs Layer 3 forwarding and sends the packet to host A through the super VLAN.

10.2.5 QinQ VLAN and Stacking VLAN

Introduction to QinQ

As Ethernet technology is deployed in more and more carrier networks (metro Ethernet networks), the standard VLAN defined in IEEE 802.1Q alone cannot completely distinguish between and isolate a large number of users. This is because the 12-bit VLAN tag field identifies a maximum of only 4096 VLANs. The 802.1Q-in-802.1Q (QinQ) technology is developed to solve this problem.

The QinQ technology improves VLAN utilization by adding another 802.1Q tag to a packet that already carries an 802.1Q tag. With this technology, the system supports up to 4096 x 4096 VLANs. QinQ allows two VLAN headers to be inserted into a single frame. Therefore, packets transmitted in the backbone network have two 802.1Q tag headers: a public network VLAN tag and a private network VLAN tag.

For details about the format of 802.1Q frames encapsulated by QinQ, see **10.2.2 VLAN Basic Concepts**.

As metro Ethernet grows and a greater variety of services are deployed, there are more scenarios in which QinQ double tags can be applied.

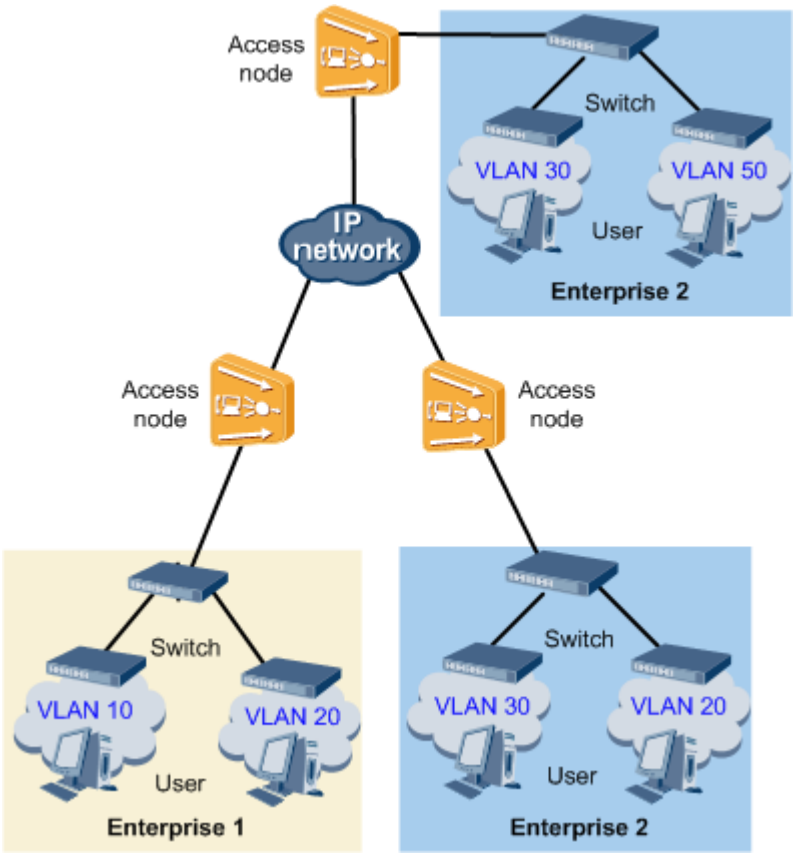
- QinQ VLAN: Packets with two VLAN tags traverse carrier networks. The inner VLAN tag is transparently transmitted to the public network. QinQ VLAN is also a simple and practical VPN technology.
- Stacking VLAN: The inner tag indicates the user; the outer tag indicates the carrier.

QinQ VLAN

In the following network, enterprise 2 has two locations, which communicate through the carrier network (public network).

 Copyright © Huawei Technologies Co., Ltd.

Figure 10-10 QinQ VLAN application



The access node provides the following configurations to deploy the QinQ VLAN so that enterprise 2 can communicate at two locations, but enterprise 1 and enterprise 2 cannot communicate with each other.

- The service VLAN is the same as the public network VLAN and the VLAN attribute is QinQ.
- The access node performs traffic classification for user packets based on service flows. Each user is mapped to one service flow. The inner and outer VLAN tags are configured for the service flows according to the VLAN plan in the following table.

Table 10-8 QinQ VLAN plan

Enterprise	Private Network VLAN (Inner VLAN)	Public Network VLAN (Outer VLAN)
Enterprise 1	VLANs 10 and 20	VLAN 10
Enterprise 2	VLANs 20, 30, and 50	VLAN 20

The following uses enterprise 2 users in the same private network (VLAN 30) as an example to describe packet processing:

1. Untagged user packets are transmitted upstream. The packets then reach the Layer 2 switch.

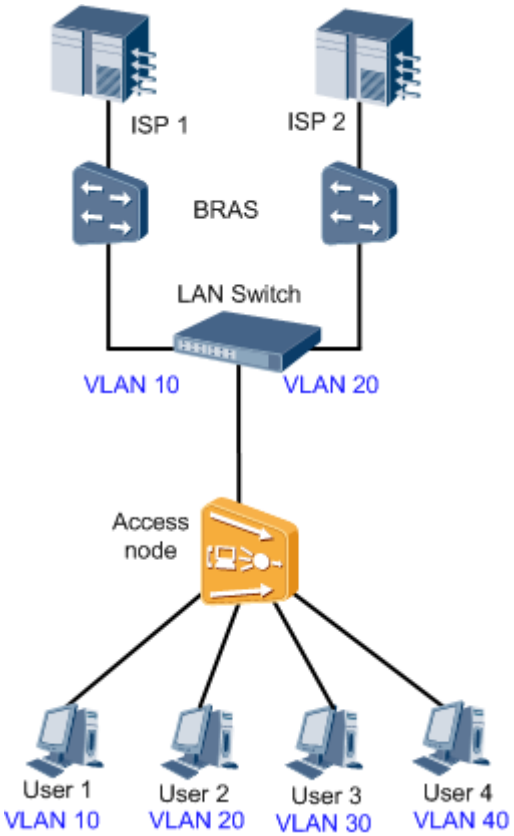
 Copyright © Huawei Technologies Co., Ltd.

2. The Layer 2 switch adds the private network VLAN tag (VLAN 30) to the packets and transmits them to the access node in the upstream direction.
3. The access node adds the public network VLAN tag (VLAN 20) to the packets and transmits the packets upstream to the upper-layer network.
4. The upper-layer network device transmits the packets to the interconnected access node according to the public network VLAN tag.
5. The interconnected access node removes the public network VLAN tag after receiving the packets, and transmits them to the switch on the same side.
6. The switch identifies and removes the private VLAN tag (VLAN 30), and forwards the untagged packets to the specified user in the private network (VLAN 30).

Stacking VLAN

The following network supports Internet service providers (ISPs) 1 and 2. User 1 and user 2 access ISP 1; user 3 and user 4 access ISP 2. The system uses the stacking VLAN to quickly provision ISP-provisioned services to specified users. Specifically, the outer VLAN tag indicates the ISP and the inner VLAN tag indicates the user.

Figure 10-11 Stacking VLAN application



The stacking VLAN is deployed according to the following configurations to connect users.

- Attribute stacking is specified for created service VLANs (different ISPs have different outer VLAN tags).

 Copyright © Huawei Technologies Co., Ltd.

- A service flow is created for each user.
- The inner and outer VLAN tags are configured according to the VLAN plan in the following table.

Table 10-9 Stacking VLAN plan

User	Inner VLAN Tag (Identifying Users)	Outer VLAN Tag (Identifying ISPs)
User 1	VLAN 10	VLAN 10
User 2	VLAN 20	
User 3	VLAN 30	VLAN 20
User 4	VLAN 40	

The following uses user 2 as an example to describe the processing of packets with the stacking VLAN from a user to an ISP:

1. User 2 sends untagged packets to the access node.
2. The access node adds two VLAN tags (inner VLAN 20 and outer VLAN 10) to the untagged packets.
3. The switch forwards the packets to ISP 1 according to outer VLAN 10.
4. ISP 1 removes outer VLAN 10 after receiving the packets and then provisions services to different users according to inner VLAN 20.

10.2.6 VLAN Translation

VLAN translation is also called VLAN mapping, which converts between user VLAN IDs and carrier VLAN IDs and allows carriers' VLAN planning to be flexible.

VLAN Translation Policies

The access node supports the following VLAN translation policies:

- N to 1 translation

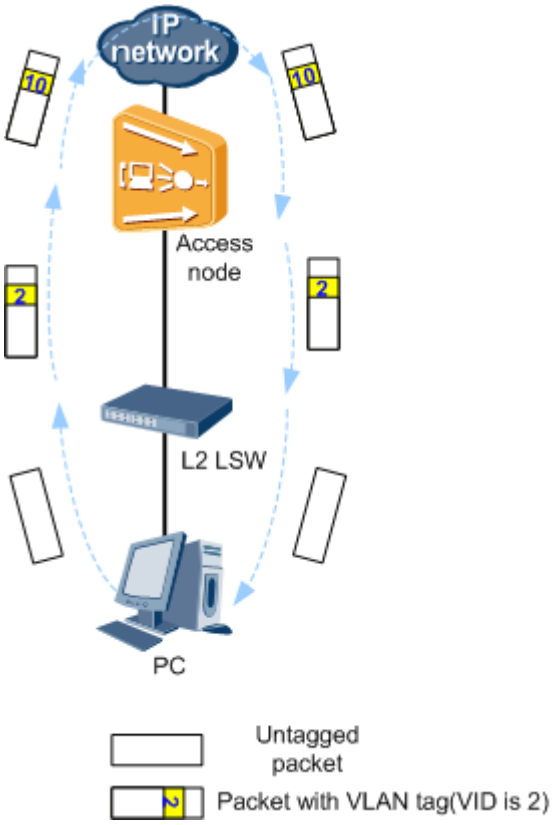
  In the receive direction, the access node translates untagged packets, single-tagged packets, and double-tagged packets into packets with a specified VLAN tag. In the transmit direction, the access node translates this specified VLAN tag to the original VLAN tag(s) when sending packets to users.
- N to 2 translation

  In the receive direction, the access node translates untagged packets, single-tagged packets, and double-tagged packets into packets with two specified VLAN tags. In the transmit direction, the access node translates these specified VLAN tags to the original VLAN tag(s) when sending packets to users.

The following figure shows the VLAN translation process, assuming that 1 to 1 VLAN translation is used. The access node translates VLAN tag (VID 2) to VLAN tag (VID 10).

 Copyright © Huawei Technologies Co., Ltd.

Figure 10-12 VLAN translation process



1. After receiving packets from the user side, the access node translates the VLAN tag of the packets according to the specified VLAN translation policy and then forwards the packets to the upper-layer network.
2. After receiving packets from the network side, the access node reverse translates the VLAN tag (restoring the original VLAN tag), and then sends the restored user packets to users.

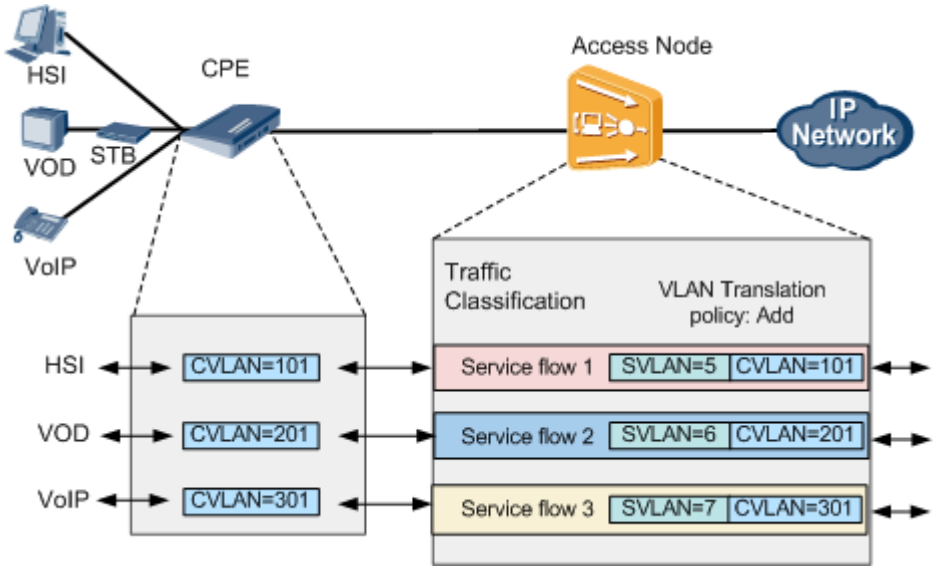
VLAN Translation Based on Service Flow

The access node performs traffic classification for user packets based on service flows. Therefore, VLAN translation configuration and function are also implemented based on service flows.

The following figure uses VLAN translation policy "add" as an example to show the VLAN translation process on the access node.

 Copyright © Huawei Technologies Co., Ltd.

Figure 10-13 VLAN translation based on service flow



1. The CPE attaches different C-VLAN tags to service packets (C-VLAN indicates the user-side VLAN, that is, the VLAN tag carried in user packets. In multi-service applications, it is usually used to identify a user).
2. The access node classifies traffic according to the C-VLAN tag and uses service flows to represent different classes.
3. The VLAN translation policy is specified to add (S-VLAN indicates the service VLAN in service flows) when service flows are configured.
4. The access node attaches an S-VLAN tag to packets matching the service flows and then forwards the packets to the upper-layer network. For example, the access node attaches VLAN tag carrying VID 5 to packets (with VLAN tag carrying VID 101) that match service flow 1.

10.2.7 VLAN Planning Suggestion

Properly plan VLAN types and attributes to meet various service requirements.

Selecting a VLAN Type

An access device connects to users in various access modes through service flows. Select a VLAN type based on service flow requirements. A standard or super VLAN does not support the creation of service flows. Therefore, a smart or MUX VLAN is used. A MUX VLAN supports only one service flow while a smart VLAN supports multiple ones. Therefore, a smart VLAN is more popular.

A standard VLAN supports the adding of only uplink ports. Therefore, it is used for managing devices.

A super VLAN applies in Layer 3 interconnection of Layer 2 isolation. Specifically, sub-VLANs in a super VLAN can communicate with each other at Layer 3 after Address Resolution Protocol (ARP) proxy is enabled.

Selecting a VLAN Attribute

Plan the number of VLAN tags for an access device based on service requirements. For example, plan one VLAN tag for simple services and two VLAN tags for

 Copyright © Huawei Technologies Co., Ltd.

wholesale services. One is used to identify Internet service providers (ISPs) and another is used to identify users. Then, select a VLAN attribute according to the number of VLAN tags.

- If one VLAN tag is required, select a common or QinQ VLAN, but not a stacking VLAN.
- If two VLAN tags are required, select a stacking or QinQ VLAN, but not a common VLAN.

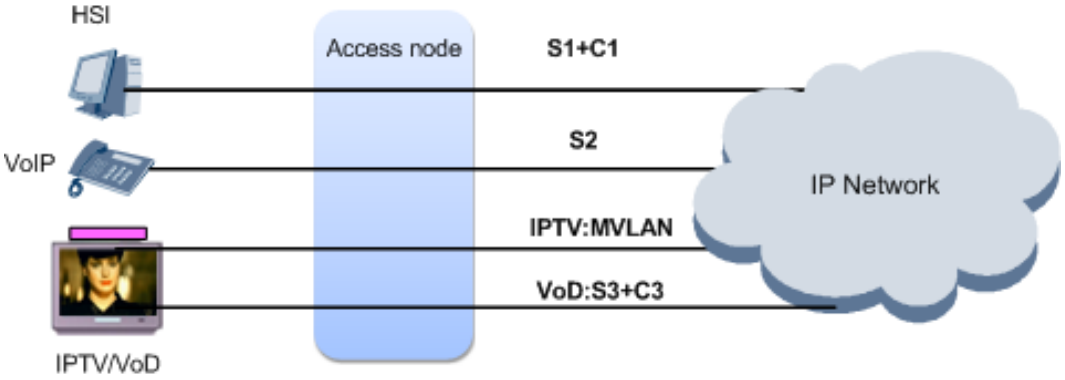
The further VLAN attribute selection rules are as follows:

- Select a QinQ VLAN for private line or transparent transmission services. The reason is that the QinQ VLAN ensures the transparent transmission of all protocol packets through hardware to the best extent.
- Select a common or stacking VLAN for enabling secure features, such as IP address anti-spoofing and MAC address anti-spoofing, Dynamic Host Configuration Protocol (DHCP) Option 82, Policy Information Transfer Protocol (PITP), and multicast functions. The reason is that a QinQ VLAN does not support the preceding functions.

VLAN Usage in the Typical Triple-play Service Scenario

In the triple-play service scenario, multiple services are provisioned for a user, such as the high-speed Internet (HSI), VoIP, IPTV, and video on demand (VoD) services.

Figure 10-14 VLAN usage in the typical triple-play service scenario



In the preceding figure:

- The HSI service uses a stacking VLAN with two VLAN tags, service VLAN (S-VLAN) tag and customer VLAN (C-VLAN) tag. S+C (S1+C1 in the preceding figure) indicates each service for each user. Enable DHCP Option 82 and PITP. Enable IP address anti-spoofing or MAC address anti-spoofing based on site requirements.
- The voice service uses a common VLAN with one VLAN tag. Enable IP address anti-spoofing and MAC address anti-spoofing.
- The VoD service uses a stacking VLAN with two VLAN tags, S-VLAN tag and C-VLAN tag (S3+C3 in the preceding figure).
- The multicast VLAN (MVLAN) of the IPTV service is a common VLAN with a VLAN tag.
- Add a VoD service flow to the MVLAN as a multicast user. The Internet Group Management Protocol (IGMP) is carried over the service flow. Therefore, the

 Copyright © Huawei Technologies Co., Ltd.

S-VLAN of the VoD service cannot be a QinQ VLAN. The reason is that a QinQ VLAN transparently transmits IGMP packets through hardware but not capture the packets. For details, see **10.2.2 VLAN Basic Concepts**.

VLAN Usage in the Typical Enterprise Private Line Service Scenario

In the enterprise private line service scenario, the headquarter (HQ) and branches of an enterprise need to communicate with each other. Both the HQ and branches are in X1 VLAN.

Figure 10-15 VLAN usage in the typical enterprise private line service scenario

In the preceding figure, the VLAN tag of both access devices 1 and 2 is S1 VLAN. The X1 VLAN transparently transmits packets. DHCP or other protocols may be used in the enterprise. Therefore, use a QinQ VLAN (S1+X1 in the preceding figure) to transparently transmit the protocol packets through hardware to the best extent. This prevents global capturing and forwarding of the protocol packets through software by mistake and ensures the performance of the enterprise private line service.

10.2.8 VLAN Translation Policies Specifications

VLAN translation policies

Table 10-10 VLAN translation policies

VLAN Tag Before Translation	VLAN Tag After Translation	VLAN Translation Policy	Description	Usage Scenario
Untagged	Service VLAN (S-VLAN)	add	An S-VLAN tag is attached.	Service flows whose S-VLAN is a QinQ VLAN Transparent LAN service (TLS) flows

Copyright © Huawei Technologies Co., Ltd.

VLAN Tag Before Translation	VLAN Tag After Translation	VLAN Translation Policy	Description	Usage Scenario
	Service VLAN +customer VLAN (S-VLAN +C-VLAN)	add double	Two VLAN tags, the outer S-VLAN tag and inner C-VLAN tag, are attached.	Single service flows whose S-VLAN is a stacking VLAN, and multiple service flows classified by the following rules. ● untagged user-side VLAN ● user-side 802.1p priority ● user-side encapsulation mode Service flows whose S-VLAN is a QinQ VLAN
C-VLAN	C-VLAN	transparent	The VLAN tag is transparently transmitted.	Service flows whose GEM port is configured with the cascade attribute Service flows whose S-VLAN is a QinQ VLAN
	S-VLAN	translate	One VLAN tag is translated.	Multiple service flows whose S-VLAN is a common VLAN, classified by the following rules. ● user-side VLAN ● combination of user-side VLAN, user-side encapsulation mode ● combination of user-side VLAN, user-side 802.1p priority Service flows whose S-VLAN is a QinQ VLAN
	S-VLAN +C-VLAN	add	An S-VLAN tag is attached.	-

 Copyright © Huawei Technologies Co., Ltd.

VLAN Tag Before Translation	VLAN Tag After Translation	VLAN Translation Policy	Description	Usage Scenario
	S-VLAN +C'-VLAN	translate and add	The C-VLAN tag is translated and the S-VLAN tag is attached.	Multiple service flows whose S-VLAN is a stacking VLAN, classified by the following rules. ● user-side VLAN ● combination of user-side VLAN, user-side encapsulation mode ● combination of user-side VLAN, user-side 802.1p priority Service flows whose S-VLAN is a QinQ VLAN
	Untagged	remove	This is a special VLAN translation mode for connection-oriented service flows. In this mode, service flows carry one VLAN tag when arriving from the destination end, and have this VLAN tag removed when transmitted from the source end. Therefore, service flows are finally untagged.	-
S-VLAN +C-VLAN	S'-VLAN +C'-VLAN	translate double	Both the S-VLAN and C-VLAN tags are translated.	Service flows whose S-VLAN is a stacking VLAN or QinQ VLAN, classified by S'-VLAN+C'-VLAN
	S'-VLAN	translate and remove	The C-VLAN tag is removed and the S-VLAN tag is translated.	Service flows whose S-VLAN is a common or QinQ VLAN, classified by S-VLAN+C-VLAN

 Copyright © Huawei Technologies Co., Ltd.

VLAN Tag Before Translation	VLAN Tag After Translation	VLAN Translation Policy	Description	Usage Scenario
	Untagged	remove double	This is a special VLAN translation mode for connection-oriented service flows. In this mode, service flows carry two VLAN tags (S-VLAN+C-VLAN) when arriving from the destination end, and have these VLAN tags removed when transmitted from the source end. Therefore, service flows are finally untagged.	-

10.2.9 Configuring a VLAN

A virtual local area network (VLAN) is used to separate broadcast domains. VLANs have enhanced security and support expansion and flexible networking. Configuring VLAN is a prerequisite for configuring a service. Hence, before configuring a service, make sure that the VLAN configuration based on planning is complete.

Application Context

VLAN application is specific to user types. For details on the VLAN application, see **Table 10-11**.

 Copyright © Huawei Technologies Co., Ltd.

Table 10-11 VLAN application and planning

Application Scenario	VLAN Planning	Remarks
The device communicates with a network management system (NMS) or the devices are cascaded.	VLAN type: standard VLAN VLAN attribute: common VLAN forwarding mode: VLAN +MAC	This plan is only applicable to Ethernet ports. A standard VLAN only contains multiple uplink ports. The Ethernet ports in one VLAN can communicate with each other and the Ethernet ports in different VLANs are separated.
N:1 access	VLAN type: smart VLAN attribute: common VLAN forwarding mode: by VLAN+MAC	In the upstream direction of a single VLAN, services of multiple users are aggregated to the same VLAN, reducing the number of VLANs. One smart VLAN can contain multiple upstream ports and multiple service ports. The service ports included in the same Smart VLAN are isolated from each other. Service ports of different VLANs are also isolated from each other.
1:1 access • Multi-ISP wholesale service (ISP is the abbreviation of Internet service provider.) • VLAN ID expansion service	VLAN type: smart Attribute: stacking VLAN forwarding mode: by S+C	A stacking VLAN packet contains two VLAN tags (inner and outer VLAN) assigned by the MA5800. The outer VLAN identifies a service (service VLAN) and the inner VLAN identifies a user (customer VLAN). The service used by a user is identified by a unique VLAN ID (S+C VLAN). The upper layer broadband remote access server (BRAS) authenticates a service by the two VLAN tags contained in the service packets. In this manner, the BRAS can receive packets from more users. On a Layer-2 upper layer network of the MA5800, packets can also be forwarded by the outer VLAN+MAC ID. This function enables ISPs to provide the wholesale services.

 Copyright © Huawei Technologies Co., Ltd.

Application Scenario	VLAN Planning	Remarks
Enterprise private line service access	VLAN type: smart VLAN attribute: QinQ VLAN forwarding mode: by VLAN+MAC or S+C.	A QinQ VLAN contains an inner VLAN tag from the user private network, and an outer VLAN tag assigned by the MA5800. The outer VLAN forms a Layer 2 virtual private network (VPN) on the user private network for transparently transmitting user services on the private network.

10.2.9.1 Creating a VLAN

Creating VLAN is a prerequisite for configuring a VLAN or service. Hence, before configuring a VLAN, make sure that the VLAN creating based on planning is complete.

Default Configuration

Table 10-12 lists the default parameter settings of VLAN.

Table 10-12 Default parameter settings of VLAN

Parameter	Default Setting	Remarks
Default VLAN of the system	VLAN ID: 1 Type: smart VLAN	You can run the **defaultvlan modify** command to modify the VLAN type but cannot delete the VLAN.
Reserved VLAN of the system	VLAN ID range: 4079-4093	You can run the **vlan reserve** command to modify the VLAN reserved by the system.

Prerequisite

- The VLAN to be added should not exist in the system.
- Service VLAN cannot be reserve VLAN.

Procedure

Step 1 Create a VLAN.

Run the **vlan** to create a VLAN. VLANs of different types are applicable to different scenarios.

 Copyright © Huawei Technologies Co., Ltd.

Table 10-13 VLAN types and application scenarios

VLAN Type	Configuration Command	VLAN Description	Application Scenario
Standard VLAN	To add a standard VLAN, run the **vlan vlanid standard** command.	Standard VLAN. Ethernet ports in a standard VLAN are interconnected with each other but Ethernet ports in different standard VLANs are isolated from each other.	Only available to Ethernet ports and specifically to network management and subtending.
Smart VLAN	To add a smart VLAN, run the **vlan vlanid smart** command.	A VLAN contains one or more xPON service ports. Service ports in the same smart VLAN are isolated. Service ports in different VLANs are also isolated. A VLAN provides access for multiple users and therefore saves VLAN resources.	Smart VLANs can be applied in residential communities to provide xPON service access.
MUX VLAN	To add a MUX VLAN, run the **vlan vlanid mux** command.	One VLAN contains only one xPON service port. The traffic streams in different VLANs are isolated from each other. One-to-one mapping can be set up between a VLAN and an access user. Therefore, a VLAN uniquely identifies an access user.	MUX VLANs are applicable to xPON service access and can distinguish users.

 Copyright © Huawei Technologies Co., Ltd.

VLAN Type	Configuration Command	VLAN Description	Application Scenario
Super VLAN	To add a super VLAN, run the **vlan vlanid super** command.	The super VLAN is based on Layer 3. One super VLAN contains multiple sub-VLANs. Through an ARP proxy, the sub-VLANs in a super VLAN can be interconnected at Layer 3.	Super VLANs save IP addresses and improve the utilization of IP addresses. For a super VLAN, sub-VLANs must be configured. You can run the **supervlan** command to add a sub-VLAN to a specified super VLAN. A sub-VLAN must be a smart VLAN or MUX VLAN.

NOTE

- To add VLANs with consecutive IDs in batches, run the **vlan vlanid to end-vlanid** command.
- To add VLANs with non-consecutive IDs in batches, run the **vlan vlan-list** command.

----End

Example

Create VLAN 50 for extension of the VLAN ID. And the type of VLAN is smart.

```
huawei(config)#vlan 50 smart
```

Create VLAN 55-60 for extension of the VLAN ID. And the type of VLAN is smart.

```
huawei(config)#vlan 55 to 60 smart
```

Create VLAN 65, 73 and 52 for extension of the VLAN ID. And the type of VLAN is smart.

```
huawei(config)#vlan 65,73,52 smart
```

10.2.9.2 Configuring the VLAN attribute

Configuring the VLAN attribute is a prerequisite for configuring a VLAN. Hence, before configuring a service, make sure that the VLAN configuration based on planning is complete.

Default Configuration

Table 10-14 lists the default parameter settings of VLAN.

Copyright © Huawei Technologies Co., Ltd.

Table 10-14 Default attribute settings of VLAN

Parameter	Default Setting
Default attribute of a new VLAN	Common

Prerequisite

- The VLAN to be configured should have been created.
- The VLAN attribute must be planned properly according to the application scenarios.

Procedure

Step 1 Configure the VLAN attribute.

The default attribute for a new VLAN is "common". You can run the **vlan attrib** command to configure the attribute of the VLAN.

Configure the attribute according to VLAN planning.

Table 10-15 VLAN attributes and application scenarios

VLAN Attribute	Configuration Command	VLAN Type	VLAN Description	Application Scenario
Common	The default attribute for a new VLAN is "common".	The VLAN with this attribute can be a standard VLAN, smart VLAN, MUX VLAN, or super VLAN.	A VLAN with the common attribute can function as a common Layer 2 VLAN or function for creating a Layer 3 interface.	Applicable to the N:1 access scenario.

 Copyright © Huawei Technologies Co., Ltd.

VLAN Attribute	Configuration Command	VLAN Type	VLAN Description	Application Scenario
QinQ VLAN	To configure QinQ as the attribute of a VLAN, run the **vlan attrib vlanid q-in-q** command.	The VLAN with this attribute can be a standard VLAN, smart VLAN or MUX VLAN. The attribute of a sub VLAN, the VLAN with a Layer 3 interface, and the default VLAN of the system cannot be set to QinQ VLAN.	The packets from a QinQ VLAN contain two VLAN tags, that is, inner VLAN tag from the private network and outer VLAN tag from the MA5800. Through the outer VLAN, a Layer 2 VPN tunnel can be set up to transparently transmit the services between private networks.	Applicable to the enterprise private line scenario. A QinQ VLAN does not support the following services: DHCP Option 82, DHCP Layer 2 relay, DHCP Layer 3 relay, PITP P mode, IP address anti-spoofing, MAC address anti-spoofing, multicast IPTV, and ARP proxy.

 Copyright © Huawei Technologies Co., Ltd.

VLAN Attribute	Configuration Command	VLAN Type	VLAN Description	Application Scenario
VLAN Stacking	To configure stacking as the attribute of a VLAN, run the **vlan attrib vlanid stacking** command.	The VLAN with this attribute can only be a smart VLAN or MUX VLAN. The attribute of a sub VLAN, the VLAN with a Layer 3 interface, and the default VLAN of the system cannot be set to VLAN stacking.	The packets from a stacking VLAN contain two VLAN tags, that is, inner VLAN tag and outer VLAN tag from the MA5800. The upper-layer BRAS authenticates the access users according to the two VLAN tags. In this manner, the number of access users is increased. On the upper-layer network in the Layer 2 working mode, a packet can be forwarded directly by the outer VLAN tag and MAC address mode to provide the wholesale service for ISPs.	Applicable to the 1:1 access scenario for the wholesale service or extension of VLAN IDs. In the case of a stacking VLAN, to configure the inner tag of the service port, run the **stacking label** command.

NOTE

- To configure attributes for the VLANs with consecutive IDs in batches, run the **vlan attrib vlanid to end-vlanid** command.
- To configure attributes for the VLANs with non-consecutive IDs in batches, run the **vlan attrib vlan-list** command.

----End

Example

To configure the attribute of VLAN 50 to **stacking** for extending VLAN IDs, do as follows:

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#vlan attrib 50 stacking
```

To configure the attributes of VLANs 55-60 (used for enterprise users) to **QinQ**, do as follows:

```
huawei(config)#vlan attrib 55 to 60 q-in-q
```

To configure the attributes of service VLANs 65, 73, and 52 to **stacking**, do as follows:

```
huawei(config)#vlan attrib 65,73,52 stacking
```

10.2.9.3 Configuring Upstream Port of a VLAN

The upstream port on an access device connects to the upper layer device to forward access device data to the upstream network and forward upper layer device data to users.

Prerequisites

The planned virtual local area network (VLAN) is already configured.

Procedure

Step 1 Configure an upstream port for the VLAN.

Run the **port vlan** command to add the upstream ports to the VLAN.

----End

Example

Assume that the 0/9/0 and 0/9/1 upstream ports are to be added to VLAN 50. To configure such upstream ports, do as follows:

```
huawei(config)#port vlan 50 0/9 0
huawei(config)#port vlan 50 0/9 1
```

10.2.9.4 Configuring a VLAN Service Profile

A VLAN service profile is a collection of service-related parameters for VLAN attributes. After a VLAN is bound to a VLAN service profile, the VLAN has all the VLAN attributes defined in the VLAN service profile. Binding a VLAN service profile is an efficient way of configuring a VLAN.

Application Context

VLAN, as a basic and also important concept of access equipment, involves discrete configurations of many parameters. These parameters include forwarding mode, security feature, protocol enabling/disabling, transparent transmission of protocol packets, and packet forwarding policy. Service parameters are related to specific VLANs and in actual usage there are a lot of VLANs, causing complex configuration. Against this backdrop, the VLAN service profile is introduced to achieve simplified and highly-efficient configuration. A VLAN service profile is abstracted from specific VLANs and supports centralized configuration of VLAN-related service parameters. Different VLANs of the same attribute can flexibly be

 Copyright © Huawei Technologies Co., Ltd.

bound to (or unbound from) a VLAN service profile to possess (or release) the attributes defined in the VLAN service profile.

Prerequisite

The VLAN to which the VLAN service profile is bound must be created.

- When a VLAN profile is bound to a QinQ VLAN, both the anti-IP spoofing and anti-MAC spoofing functions are unavailable.
- SVLAN+CVLAN forwarding and anti-MAC spoofing cannot be enabled at the same time in a VLAN profile.
- After a VLAN profile is bound to a VLAN, if the VLAN attribute is Common, the packet forwarding mode cannot be set to vlan-connect (SVLAN+CVLAN forwarding).

Configuration Process

1. Create a VLAN service profile.
2. Configure the following service parameters according to service requirements:
   - Forwarding mode
   - Forwarding policy
   - Protocol switch
   - Transparent transmission function
   - Security function
3. Commit to save the current parameters.
4. Bind the VLAN service profile to the VLAN.

Procedure

Step 1 Create a VLAN service profile.

Run the **vlan service-profile** command to create a VLAN service profile or enter the configuration mode of the VLAN service profile. When the profile does not exist, running this command to create a VLAN service profile and enter the configuration mode of the service profile. If the profile exists, running this command to directly enter the configuration mode of this service profile.

Step 2 Configure the VLAN forwarding mode.

Forwarding mode refers to the Layer 2 packet forwarding mechanism, including VLAN+MAC forwarding (default) and SVLAN+CVLAN (or S+C) forwarding. In VLAN+MAC forwarding, the system needs to dynamically learn the mapping relationship between VLAN, source MAC address, and port. In S+C forwarding, the system does not need to dynamically learn MAC addresses but determines the forwarding entry according to SVLAN and CVLAN. Because S+C forwarding does not depend on MAC address learning, it has the following advantages:

1. Saving MAC addresses
2. Preventing occurrence of unknown unicast packets caused by aging of dynamic MAC addresses Broadcasting unknown unicast packets threatens the security of the device

 Copyright © Huawei Technologies Co., Ltd.

3. Ensuring security by solving problems such as MAC spoofing and attack

- Run the **forwarding** command to configure the VLAN forwarding policy.
- Run the **user-bridging** command to configure the bridging function of the VLAN service profile. After the bridging function is enabled, two users in the same VLAN can directly communicate with each other at Layer 2.

Step 3 Configure the VLAN forwarding policy.

Forwarding policy refers to the discard policy of packets such as downstream broadcast, downstream unknown unicast, and unknown multicast packets.

- Run the **packet-policy** command to configure the forwarding policy for the downstream broadcast packets, downstream unknown unicast packets, and unknown multicast packets in the VLAN. Two policies namely forward and discard are supported.
- Run the **igmp mismatch** command to configure the mismatch IGMP policy of the VLAN, supports the transparent and discard policies.

Step 4 Configure the VLAN protocol switch.

Protocol switch refers to whether to enable certain types of protocols or certain functions of a protocol. VMAC aging and PPPoE MAC conflict with S+C forwarding.

- Run the **dhcp mode** command to configure the DHCPv4 forwarding mode, that is, to switch between the DHCP Layer 2 forwarding mode and the DHCP Layer 3 forwarding mode.
- Run the **dhcpv6 mode** command to configure the DHCPv6 forwarding mode, that is, to switch between the DHCP Layer 2 forwarding mode and the DHCP Layer 3 forwarding mode.
- Run the **pppoe mac-mode** command to configure the MAC address allocation mode for PPPoE users. Two modes namely, single-mac and multi-mac are supported.
- Run the **vmac aging-mode** command to configure the VMAC aging mode, which can be common aging or DHCP-based aging.
- Run the **pitp** command to configure the PITP function to implement authentication of bound user account and access port.
- Run the **dhcp option82** command to configure the DHCPv4 option 82 feature.
- Run the **dhcpv6 option82** command to enable or disable the DHCPv6 option feature.
- Run the **dhcp proxy** command to configure the DHCP proxy function. After the DHCP proxy function is enabled, the server ID proxy function and lease time proxy function will be enabled.

Step 5 Configure the VLAN transparent transmission function.

In transparent transmission, the system does not process specified types of protocol packets but transparently transmits them.

- Run the **bpdu tunnel** command to configure the BPDU transparent transmission switch. After transparent transmission is enabled, the Layer 2 BPDUs of the private network can be transmitted transparently over the public network.

 Copyright © Huawei Technologies Co., Ltd.

- Run the **vtp-cdp tunnel** command to configure the VTP/CDP packet transparent transmission switch. After the switch is enabled, VTP/CDP packets are transparently transmitted based on the VLAN.
- Run the **rip tunnel** command to configure the RIP Layer 2 transparent transmission switch. After the transparent transmission switch is enabled, RIP packets can be transparently transmitted at Layer 2 based on VLAN on the device without running the RIP protocol.
- Run the **l3-protocol tunnel** command to configure the L2 transparent transmission for the L3 protocol packets except RIP and OSPF packets. After this function is enabled, the L3 protocol packets can be transparently transmitted at Layer 2 based on VLAN on the device without running the L3 protocol.
- Run the **ipv6 dad proxy** command to configure the DAD proxy (duplicate address detect proxy). DAD proxy prevents repeated LLA configuration on the user side.

Step 6 Configure the VLAN security function.

The security function is used to prevent malicious users from attacking the system by forging the IP address or MAC address of an authorized user. VMAC and anti-MAC spoofing conflict with S+C forwarding.

- Run the **security anti-ipspoofing** command to configure the anti-IPv4 spoofing function. After the anti-IPv4 spoofing function is enabled, the system automatically and dynamically binds the IPv4 address to the user. The packet can be transmitted upstream through the device only when the source IPv4 address of the packet is the same as the bound IPv4 address. Otherwise, the packet is discarded.
- Run the **security anti-ipv6spoofing** command to configure the anti-IPv6 spoofing function. After the anti-IPv6 spoofing function is enabled, the system automatically and dynamically binds the IPv6 address to the user. The packet can be transmitted upstream through the device only when the source IPv6 address of the packet is the same as the bound IPv6 address. Otherwise, the packet is discarded.
- Run the **security anti-macspoofing** command to configure the anti-MAC spoofing function. After the anti-MAC spoofing function is enabled, the system automatically and dynamically binds the MAC address to the traffic stream. When the source MAC address of the traffic stream is the same as the bound MAC address, the traffic stream can be upstream transmitted through the device. Otherwise, the packets are discarded.
- Run the **security arp-reply** command to enable the network-side ARP proxy response function. If the network-side ARP proxy response function is enabled, the system searches for user's going online information based on the destination IP address and VLAN after it receives network-side ARP request packets. If there is an online user, the system performs proxy response. If there is no online user, the system discards or forwards the ARP request packets based on the setting in the **security arp-reply unknown-policy** command. This prevents ARP request packets from being sent to user ports and reduces system resources.
- Run the **security ns-reply** command to enable the network-side NS proxy response function. If network-side NS proxy reply is enabled, the system searches for user's going online information based on the destination IP address and VLAN after it receives network-side NS packets. If there is an

 Copyright © Huawei Technologies Co., Ltd.

online user, the system performs proxy response. If there is no online user, the system discards or forwards the NS packets based on the setting in the **security ns-reply unknown-policy** command. This prevents NS packets from being sent to user ports and reduces system resources.

- Run the **security bind-route-nd** command to configure the function of binding route with neighbor entry. After the function of binding route with neighbor entry is enabled, the system automatically generates the route and neighbor entry of a DHCPv6 user based on the user information recorded when the user goes online. This function reduces the effort of configuring static routes manually, prevents neighbor packets from being sent to the user side, and enhances system security.
- Run the **vmac** command to enable or disable VMAC. By default, VMAC is disabled.

Step 7 Commit to save the current parameters.

Run the **commit** command to commit the current parameter configuration of the VLAN service profile. After the configuration is completed, run the commit command to make the configuration take effect.

Step 8 Bind the VLAN service profile to the VLAN.

Run the **vlan bind service-profile** command to bind the configured VLAN service profile to a specified VLAN. After the binding, the VLAN-level feature control switch is based on the configuration of the VLAN service profile. Independent configuration commands for VLAN-based features are no longer effective.

----End

Result

You can query the configuration of the VLAN service profile by the **display vlan service-profile** command.

After a VLAN service profile is bound to a VLAN, regarding the parameters whose **Committed** state is **NotConfig**, the configuration commands that are independent of the VLAN take effect; other parameters adopt the control parameters of the profile. Modifying the feature parameters relevant to the VLAN does not take effect.

Example

Add VLAN service profile 3 and bind it to VLAN 100. The profile parameters are planned as follows:

- VLAN forwarding mode VLAN+MAC address (vlan-mac)
- BPDU transparent transmission: enabled
- Unknown multicast packet: discarded

Adopt the default values for other parameters.

```
huawei(config)#vlan service-profile profile-id 3
huawei(config-vlan-srvprof-3)#forwarding vlan-mac
huawei(config-vlan-srvprof-3)#bpdu tunnel enable
huawei(config-vlan-srvprof-3)#packet-policy multicast discard
huawei(config-vlan-srvprof-3)#commit
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-vlan-srvprof-3)#quit
huawei(config)#vlan bind service-profile 100 profile-id 3
```

10.2.9.5 Example: Configuring the QinQ VLAN

The QinQ-VLAN-based private line service can achieve the interconnection and secure communication among branches in different areas within the enterprise private network.

Prerequisites

- Network devices and lines must be in the normal state.
- The authentication data of the access user must be configured on the BRAS.
- The system is working properly.

Service Requirements

- An enterprise requires to achieve the interconnection and secure communication between its headquarters and the branches located in different areas through Layer 2 switching network, and to isolate the data of different departments.
- The access device uses PON access.

Networking

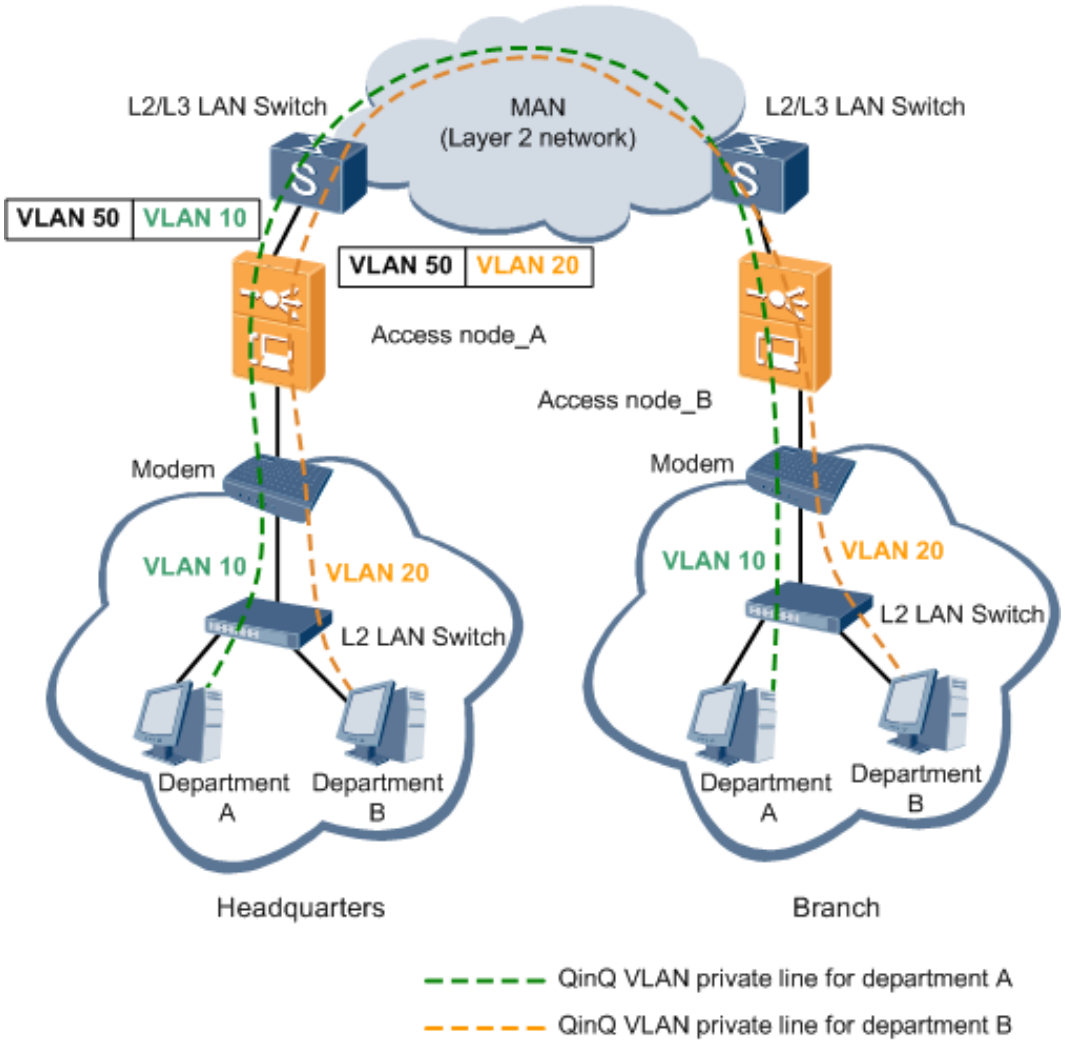
Figure 10-16 shows an example network for configuring the private line service.

The two branches of the enterprise are connected to the (metropolitan area network) MAN through the MA5800. The upper-layer network must work in the Layer 2 mode, and must forward packets according to the VLAN and the MAC address.

On the MA5800, the attribute of the upstream VLAN of user packets is configured as QinQ private line service. A VLAN tunnel is created in Layer 2/Layer 3 MAN for transmitting data carrying the VLAN tag. Different VLAN IDs are used for different departments to achieve user isolation and data security. In this way, the service packets of the enterprise private network can be transparently transmitted through the public network, and the two branches can communicate with each other securely.

 Copyright © Huawei Technologies Co., Ltd.

Figure 10-16 Example network for configuring the private line service



The configuration on MA5800_A is the same as the configuration on MA5800_B. The following uses the configuration on MA5800_A as examples to describe how to configure the private line service implemented through a QinQ VLAN.

Data Plan

Table 10-16 lists the key data planning of the QinQ VLAN private line service on MA5800_A.

 Copyright © Huawei Technologies Co., Ltd.

Table 10-16 Data plan for configuring the QinQ VLAN private line service

Item	Data	Remarks
VLAN	• VLAN ID: 50 • VLAN type: smart VLAN • VLAN attribute: QinQ	The public network VLAN, namely outer VLAN, is configured here. The inner VLAN belongs to the enterprise private network, which can be planned by the enterprise self.
Transparent transmission of Layer 2 packets	Enable the transparent transmission of BPDUs.	This function is mainly used in the QinQ service to provide a transparent and secure data channel for the enterprise branches located in two places within the enterprise private network. After the transparent transmission of BPDUs is enabled, the Layer 2 protocol packets of a private network can be transparently transmitted through the public network.
Upstream port	0/9/0	-
Traffic profile	• Traffic profile ID: 10 • CIR: 4 Mbit/s • Priority copy policy: user-cos • Default 802.1p priority of the packet: 4 • Priority-based scheduling policy of the downstream packets: local-setting	The CIR depends on the user bandwidth requirement.

 Copyright © Huawei Technologies Co., Ltd.

Item	Data	Remarks
GPON port	• GPON port ID: 0/2/0 • DBA profile: – Profile type: type3 – Profile name: DBA_bandwidth – Assured bandwidth: 10 Mbit/s • ONT line profile – Profile ID: 10 – T-CONT ID: 4 • ONT service profile ID: 10 • ONT port VLAN ID: 10 • Default ONT alarm profile ID: 1	-
Service virtual port	• Service port: 0/2/0 • ID of the traffic profile in Rx direction: 10 • ID of the traffic profile in Tx direction: 10	• The traffic profile that meets the service requirement is used.

Procedure

Step 1 Configure a DBA profile.

The profile type is Type3, assured bandwidth is 10 Mbit/s, and profile name is DBA_10M.

```
huawei(config)#dba-profile add profile-name DBA_10M type3 assure 10240 max 10240
```

Step 2 Configure a GPON ONT line profile.

Create a GPON ONT line profile with an index number of 10, and T-CONT ID is 4.

```
huawei(config)#ont-lineprofile gpon profile-id 10
huawei(config-gpon-lineprofile-10)#tcont 4 dba-profile-name DBA_10M
huawei(config-gpon-lineprofile-10)#gem add 10 eth tcont 4
huawei(config-gpon-lineprofile-10)#gem mapping 10 0 vlan 10
huawei(config-gpon-lineprofile-10)#commit
huawei(config-gpon-lineprofile-10)#quit
```

Step 3 Configure a GPON ONT service profile.

Create a GPON ONT service profile with index 10, set the VLAN ID of the ONT port to 10, and set the port capability set of the ONT to adaptive. The system automatically adapts to the actual capability of online ONTs.

```
huawei(config)#ont-srvprofile gpon profile-id 10
huawei(config-gpon-srvprofile-10)#ont-port eth adaptive
huawei(config-gpon-srvprofile-10)#port vlan eth 1 10
huawei(config-gpon-srvprofile-10)#commit
huawei(config-gpon-srvprofile-10)#quit
```

Step 4 Configure a GPON ONT.

 Copyright © Huawei Technologies Co., Ltd.

Enable the automatic ONT discovery function. The following uses port 0/2/0 as an example. The configuration of the other GPON ports is similar to that of port 0/2/0.

NOTE

- By default, a port is in the activated state. Before binding a profile to the port, you must deactivate the port.

```
huawei(config)#interface gpon 0/2
huawei(config-if-gpon-0/2)#port 0 ont-auto-find enable
huawei(config-if-gpon-0/2)#ont add 0 password-auth 0100000001 always-on omci ont-lineprofile-id 10 ont-srvprofile-id 10
huawei(config-if-gpon-0/2)#quit
```

Step 5 Create a VLAN.

The VLAN ID is 50, and the VLAN is a smart VLAN.

```
huawei(config)#vlan 50 smart
```

Step 6 Set the VLAN attribute to QinQ.

```
huawei(config)#vlan attrib 50 q-in-q
```

Step 7 Enable the transparent transmission of BPDUs.

Enable the transparent transmission of BPDUs so that the Layer 2 protocol packets of a private network can be transparently transmitted through the public network. By default, the transparent transmission of BPDUs is disabled.

```
huawei(config)#bpdu tunnel vlan 50 enable
```

Step 8 Add an upstream port to the VLAN.

Add upstream port 0/9/0 to VLAN 50.

```
huawei(config)#port vlan 50 0/9 0
```

Step 9 Add a traffic profile.

The profile ID is 10, the CIR is 4 Mbit/s, and packets are scheduled according to the priority specified in the traffic profile.

```
huawei(config)#traffic table ip index 10 cir 4096 priority user-cos 4 priority-policy local-Setting
```

Step 10 Add a service port to the VLAN.

Add a service port to the VLAN 50, and use traffic profile 10 that meets the service requirements. The user port is 0/2/0.

```
huawei(config)#service-port vlan 50 gpon 0/2/0 ont 1 gemport 10 inbound traffic-table index 10 outbound traffic-table index 10
```

Step 11 Save the data.

```
huawei(config)#save
```

----End

Result

After the configuration, the two branches of the enterprise can communicate with each other.

 Copyright © Huawei Technologies Co., Ltd.

Configuration File

```
dba-profile add profile-name DBA_10M type3 assure 10240 max 10240
ont-lineprofile gpon profile-id 10
tcont 4 dba-profile-name DBA_10M
gem add 10 eth tcont 4
gem mapping 10 0 vlan 10
commit
quit
ont-srvprofile gpon profile-id 10
ont-port eth adaptive
port vlan eth 1 10
commit
interface gpon 0/2
port 0 ont-auto-find enable
ont add 0 password-auth 0100000001 always-on omci ont-lineprofile-id 10 ont-srvprofile-id 10
quit
vlan 50 smart
vlan attrib 50 q-in-q
bpdu tunnel vlan 50 enable
port vlan 50 0/9 0
traffic table ip index 10 cir 4096 priority user-cos 4 priority-policy local-Setting
service-port vlan 50 gpon 0/2/0 ont 1 gemport 10 inbound traffic-table index 10 outbound traffic-table
index 10
save
```

10.2.9.6 Example: Configuring VLAN Stacking Multi-ISP Wholesale Access

In a Layer-2 switching metropolitan area network (MAN), there are multiple Internet service providers (ISPs). To provision the services provided by the ISP to the specified user group rapidly, the outer VLAN tags of VLAN stacking can be used to identify ISPs, while the inner VLAN tags to identify users. In this way, different user groups can be connected to the specified ISPs in batches through different outer VLAN tags to obtain services from the ISPs.

Prerequisites

- Network devices and lines must be in the normal state.
- The authentication data of the access user must be configured on the BRAS.
- The system is working properly.

Service Requirements

- The two ISPs in the Layer-2 switching MAN provide broadband services to enterprise users. Different enterprise user groups are bulk connected to the specified ISP to obtain services provided by the ISP.
- The access device uses GPON access.

Networking

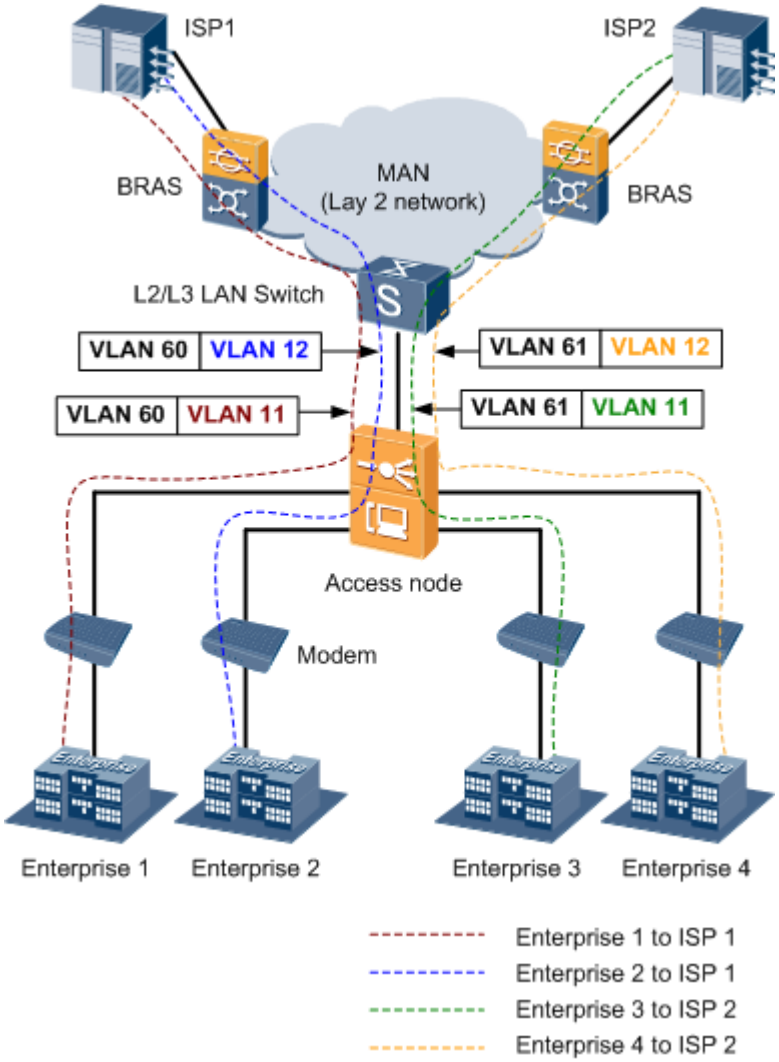
Figure 10-17 shows an example network for configuring the VLAN stacking multi-ISP wholesale access.

Enterprise 1 and 2 belong to ISP 1, and enterprise 3 and 4 belong to ISP 2. Based on the VLAN stacking feature. The upper-layer network must work in the Layer 2 mode, and must forward packets according to the VLAN and the MAC address. The MA5800 adds the outer VLAN tag to differentiate ISPs and inner VLAN tag to differentiate users and forwards the user packet to the Layer 2 network. Then the Layer 2/Layer 3 LAN switch forwards the user packets to the specified ISP BRAS

 Copyright © Huawei Technologies Co., Ltd.

based on the outer VLAN tag. The ISP BRAS removes the outer VLAN tag and identify the users based on the inner VLAN tag. After passing the authentication, the enterprise users can obtain various services provided by the ISP.

Figure 10-17 Example network for configuring the VLAN stacking multi-ISP wholesale access



Data Plan

Table 10-17 lists the key data planning of the VLAN stacking wholesale service on the MA5800.

 Copyright © Huawei Technologies Co., Ltd.

Table 10-17 Data plan for configuring the VLAN stacking wholesale service

Item	Data	Remarks
VLAN	VLAN ID: ● ISP1: 60 – Enterprise user 1: 11 – Enterprise user 2: 12 ● ISP2: 61 – Enterprise user 3: 11 – Enterprise user 4: 12 ● VLAN type: smart VLAN ● VLAN attribute: stacking	The outer VLAN identifies the ISP, and inner VLAN identifies the user. The inner VLAN tags must be unique under the same ISP, but can be the same under different ISPs.
Upstream port	0/9/0	-
Traffic profile	● Traffic profile ID: 10 ● CIR: 2 Mbit/s ● Priority copy policy: user-cos ● Default 802.1p priority of the packet: 4 ● Priority-based scheduling policy of the downstream packets: local-setting	The CIR depends on the user bandwidth requirement.
GPON port	● GPON port ID: – Enterprise user 1: 0/2/0 – Enterprise user 2: 0/2/1 – Enterprise user 3: 0/3/0 – Enterprise user 4: 0/3/1 ● DBA profile: – Profile type: type3 – Profile name: DBA_bandwidth – Assured bandwidth: 10 Mbit/s ● ONT line profile – Profile ID: 10 – T-CONT ID: 4 ● Default ONT service profile ID: 0 ● Default ONT alarm profile ID: 1	-

 Copyright © Huawei Technologies Co., Ltd.

Item	Data	Remarks
Service virtual port	• Service port ID: – Enterprise user 1: 0 – Enterprise user 2: 1 – Enterprise user 3: 2 – Enterprise user 4: 3 • ID of the traffic profile in Rx direction: 10 • ID of the traffic profile in Tx direction: 10	• The VPI/VCI is the same as the management VPI/VCI on the peer modem.

Procedure

Step 1 Configure a DBA profile.

The profile type is Type3, assured bandwidth is 10 Mbit/s, and profile name is DBA_10M.

```
huawei(config)#dba-profile add profile-name DBA_10M type3 assure 10240 max 10240
```

Step 2 Configure a GPON ONT line profile.

Create a GPON ONT line profile with an index number of 10, and T-CONT ID is 4.

```
huawei(config)#ont-lineprofile gpon profile-id 10
huawei(config-gpon-lineprofile-10)#tcont 4 dba-profile-name DBA_10M
huawei(config-gpon-lineprofile-10)#gem add 10 eth tcont 4
huawei(config-gpon-lineprofile-10)#gem add 11 eth tcont 4
huawei(config-gpon-lineprofile-10)#gem mapping 10 0 vlan 60
huawei(config-gpon-lineprofile-10)#gem mapping 11 0 vlan 61
huawei(config-gpon-lineprofile-10)#commit
huawei(config-gpon-lineprofile-10)#quit
```

Step 3 Create VLANs.

The outer VLAN IDs are 60 and 61, and the VLANs are smart VLANs.

```
huawei(config)#vlan 60-61 smart
  It will take several minutes, and console may be timeout, please use command
idle-timeout to set time limit
  Are you sure to add VLANs? (y/n)[n]:y
```

Step 4 Set the VLAN attribute to stacking.

```
huawei(config)#vlan attrib 60-61 stacking
  It will take several minutes, and console may be timeout, please use command
idle-timeout to set time limit
  Are you sure to continue? (y/n)[n]:y
```

Step 5 Add an upstream port to the VLANs.

Add an upstream port 0/9/0 to VLANs 60 and 61.

```
huawei(config)#port vlan 60-61 0/9 0
  It will take several minutes, and console may be timeout, please use command
idle-timeout to set time limit
  Are you sure to add standard port(s)? (y/n)[n]:y
```

 Copyright © Huawei Technologies Co., Ltd.

Step 6 Add a traffic profile.

The profile ID is 10, the CIR is 2 Mbit/s, and packets are scheduled according to the priority specified in the traffic profile.

```
huawei(config)#traffic table ip index 10 cir 2048 priority user-cos 4 priority-policy local-Setting
```

Step 7 Configure a GPON ONT.

Enable the automatic ONT discovery function. The following uses port 0/2/0 as an example. The configuration of the other three GPON ports are similar to that of port 0/2/0.

```
huawei(config)#interface gpon 0/2
huawei(config-if-gpon-0/2)#port 0 ont-auto-find enable
huawei(config-if-gpon-0/2)#ont add 0 password-auth 0100000001 always-on omci ont-lineprofile-id 10 ont-srvprofile-id 0
huawei(config-if-gpon-0/2)#quit
```

Step 8 Add service ports to the VLANs.

Add service ports to the VLANs, and use traffic profile 10 that meets the service requirements.

```
huawei(config)#service-port 0 vlan 60 gpon 0/2/0 ont 1 gemport 10 rx-cttr 10 tx-cttr 10
huawei(config)#service-port 1 vlan 60 gpon 0/2/1 ont 1 gemport 10 rx-cttr 10 tx-cttr 10
huawei(config)#service-port 2 vlan 61 gpon 0/3/0 ont 1 gemport 11 rx-cttr 10 tx-cttr 10
huawei(config)#service-port 3 vlan 61 gpon 0/3/1 ont 1 gemport 11 rx-cttr 10 tx-cttr 10
```

Step 9 Set the inner VLAN tags.

The inner VLAN tag identifies the user. Note that the inner VLAN tag must be unique in one ISP domain, and the inner VLAN tags can be the same in different ISP domains.

```
huawei(config)#stacking label service-port 0 11
huawei(config)#stacking label service-port 1 12
huawei(config)#stacking label service-port 2 11
huawei(config)#stacking label service-port 3 12
```

Step 10 Save the data.

```
huawei(config)#save
```

----End

Result

- After passing the authentication by the ISP1 BRAS, enterprise 1 and enterprise 2 can obtain the service provided by ISP1.
- After passing the authentication by the ISP2 BRAS, enterprise 3 and enterprise 4 can obtain the service provided by ISP2.

Configuration File

```
dba-profile add profile-name DBA_10M type3 assure 10240 max 10240
ont-lineprofile gpon profile-id 10
tcont 4 dba-profile-name DBA_10M
gem add 10 eth tcont 4
gem add 11 eth tcont 4
gem mapping 10 0 vlan 60
gem mapping 11 0 vlan 61
commit
quit
vlan 60-61 smart
```

 Copyright © Huawei Technologies Co., Ltd.

```
  It will take several minutes, and console may be timeout, please use command
idle-timeout to set time limit
  Are you sure to add VLANs? (y/n)[n]:y
  The total of the VLANs having been processed is 2
  The total of the added VLANs is 2

vlan attrib 60-61 stacking
  It will take several minutes, and console may be timeout, please use command
idle-timeout to set time limit
  Are you sure to continue? (y/n)[n]:y
  The total of the VLANs having been processed is 2
  The total of the VLAN(s) which have been operated successfully is 2
port vlan 60-61 0/9 0
    It will take several minutes, and console may be timeout, please use command
idle-timeout to set time limit
  Are you sure to add standard port(s)? (y/n)[n]:y
  The total of the VLANs having been processed is 2
  The total of the port VLAN(s) having been added is 2
traffic table ip index 100 cir 2048 priority user-cos 4 priority-policy local-Setting
interface gpon 0/2
port 0 ont-auto-find enable
ont add 0 password-auth 0100000001 always-on omci ont-lineprofile-id 0 ont-srvprofile-id 0
quit
service-port 0 vlan 60 gpon 0/2/0 ont 1 gemport 10 rx-cttr 10 tx-cttr 10
service-port 1 vlan 60 gpon 0/2/1 ont 1 gemport 10 rx-cttr 10 tx-cttr 10
service-port 2 vlan 61 gpon 0/3/0 ont 1 gemport 11 rx-cttr 10 tx-cttr 10
service-port 3 vlan 61 gpon 0/3/1 ont 1 gemport 11 rx-cttr 10 tx-cttr 10
stacking label service-port 0 11
stacking label service-port 1 12
stacking label service-port 2 11
stacking label service-port 3 12
save
```

10.2.10 VLAN Reference Standards and Protocols

The following lists the reference standards and protocols of this feature:

Table 10-18 Reference standards and protocols of VLAN feature

Document name	Description
IEEE 802.1Q	IEEE standards for Local and metropolitan area networks-Virtual Bridged Local Area Networks
IEEE 802.1ad	Virtual Bridged Local Area Networks Amendment 4: Provider Bridges
RFC3069	VLAN Aggregation for Efficient IP Address Allocation

10.3 Service Flow

10.3.1 Introduction to Service Flow

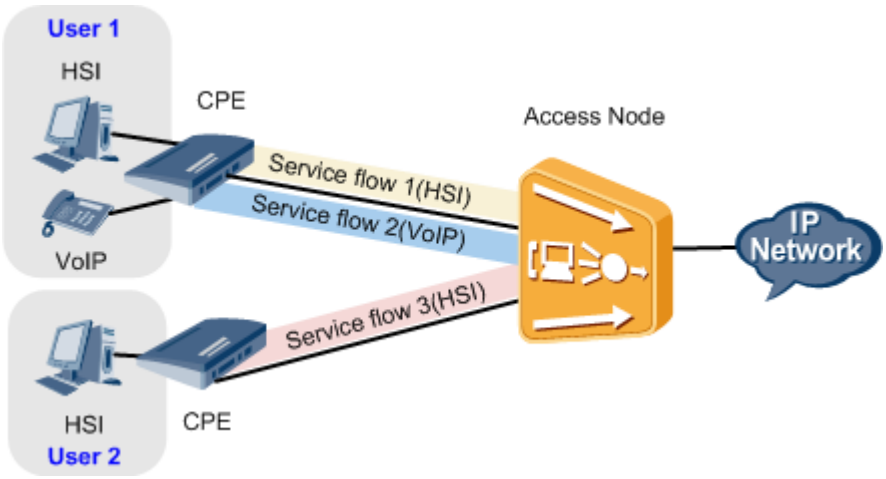
An access device provides services to a large number of users, and each user requires multiple types of services (for example, HSI and VoIP services). The access device is required to differentiate between different user services when processing user packets, so that the services do not interfere with each other. To address this requirement, the access node provides the service flow feature.

 Copyright © Huawei Technologies Co., Ltd.

Definition

Service flow, also called service port, is a result of traffic classification by characteristics of an Ethernet packet on a physical or logical port. Service flow is also a Layer 2 logical channel that carries services between the access node and users (Specify the Layer 2 forwarding path).

Figure 10-18 Service flows model



Services from different users (service flows 1 and 3) or services from the same user (service flows 1 and 2) are carried over different service flows.

Purpose

A service flow is the basis for provisioning each type of service on the access node. That is, the configuration of service flows is mandatory. In addition to traffic classification, a service flow is the smallest unit of user service processing. Hence, differentiated and fine-grained management, such as QoS, line identification, and security policies, of user services can be implemented based on service flows.

10.3.2 Service Flow Principles

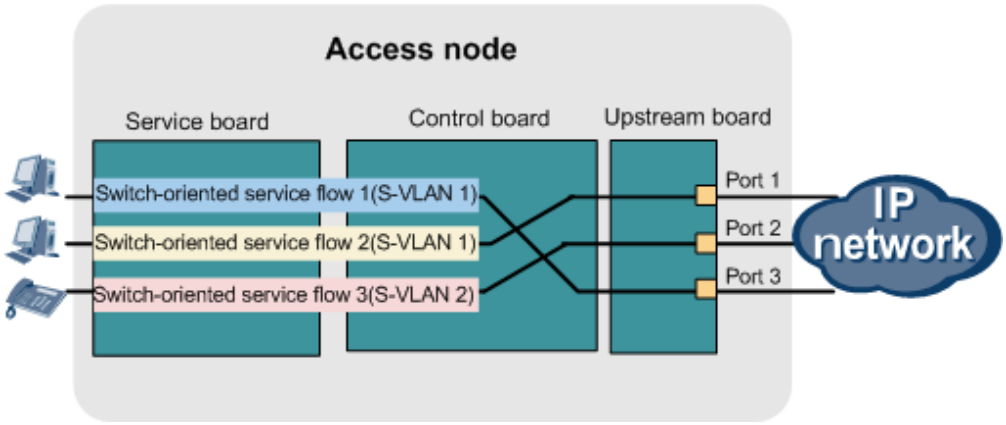
Service Flow Classification

Based on the service type, service flows can be classified as follows:

- **Switch-oriented service flow**: It is a Layer 2 service channel from the user access port on a device to the VLAN (S-VLAN) of the switching core. In an S-VLAN, one upstream port can be bound to multiple switch-oriented service flows and one switch-oriented service flow can be bound to multiple upstream ports. Since user information is terminated at service boards, S-VLAN based switching is performed on the MA5800. Therefore, the service flows transmitted in the Layer 2 service channel are called switch-oriented service flows.

 Copyright © Huawei Technologies Co., Ltd.

Figure 10-19 The schematic diagram for switch-oriented service flows



NOTE

Switching on the MA5800: When a packet arrives at the MA5800, the MA5800 selects an egress port and forwards the packet from this port. This process is called switching (or forwarding) on the MA5800.

Traffic Classification Rules

Packets are classified based on traffic classification rules, with each type of packets corresponding to a service flow. The traffic classification rule for a service flow is specified when you run the **service-port** command to add the service flow. After packets are mapped to different service flows, the system processes the service flows based on the configured policies (including VLAN translation policy and QoS policy).

Traffic classification rules supported by the MA5800 can be generalized into two types:

- Classification based on physical ports or logical ports.
- Classification based on Ethernet packet attributes (such as priority).

 Copyright © Huawei Technologies Co., Ltd.

Table 10-19 Traffic classification rules supported by the MA5800

Service Flow Type	Classification Rule Based on Physical Ports or Logical Ports		Classification Rule Based on Ethernet Packet Attributes
	Physical Port	Logical Port Identifier	
Switch-oriented service flow	Indicates the type of the physical port for a user to access the MA5800. The options are as follows: • xPON • Ethernet • Ethernet ports on ONTs in xPON access mode: The service flow created using this type of port is an end-to-end service flow between an OLT and an ONT.	• GPON access mode: GEM port • xPON access mode: IPHOST voice port on an ONT. The IPHOST voice port is a virtual port of an ONT and implements communication between the voice chip and GMAC chip. The service flow created using this port is an end-to-end (voice) service flow between an OLT and an ONT.	Switch-oriented service flows support all the classification rules Transparent LAN service (TLS) flows are also called other-all service flows. After user packets are transparently transmitted to a VLAN, VLAN tags are added to the packets. When the packets are transmitted out of the network, VLAN tags are stripped.

 Copyright © Huawei Technologies Co., Ltd.

Attributes of Service Flows

Table 10-20 Attributes of service flows

Name	Description	Configuration Command
Service flow index	Each service flow has a unique index on the MA5800. The index can be used when other features based on the service flow are configured. For example, you can configure a user as a multicast user simply by specifying the service flow index.	**service-port** *index*
Description of the service flow	There are usually a large number of service flows in an access node. The description (for example, the area in which the user is located and service characteristics) of a service flow facilitates service maintenance.	**service-port desc**
Description of the remote device of the service flow	The description includes information about the remote device of the user side of a service flow and helps determine the terminal type when many types of terminals are used, which facilitates maintenance.	**service-port remote-desc**
Administrative status	Users can pause and resume the transmission of services on a service flow by setting the administrative status of the service flow.	**service-port adminstatus**

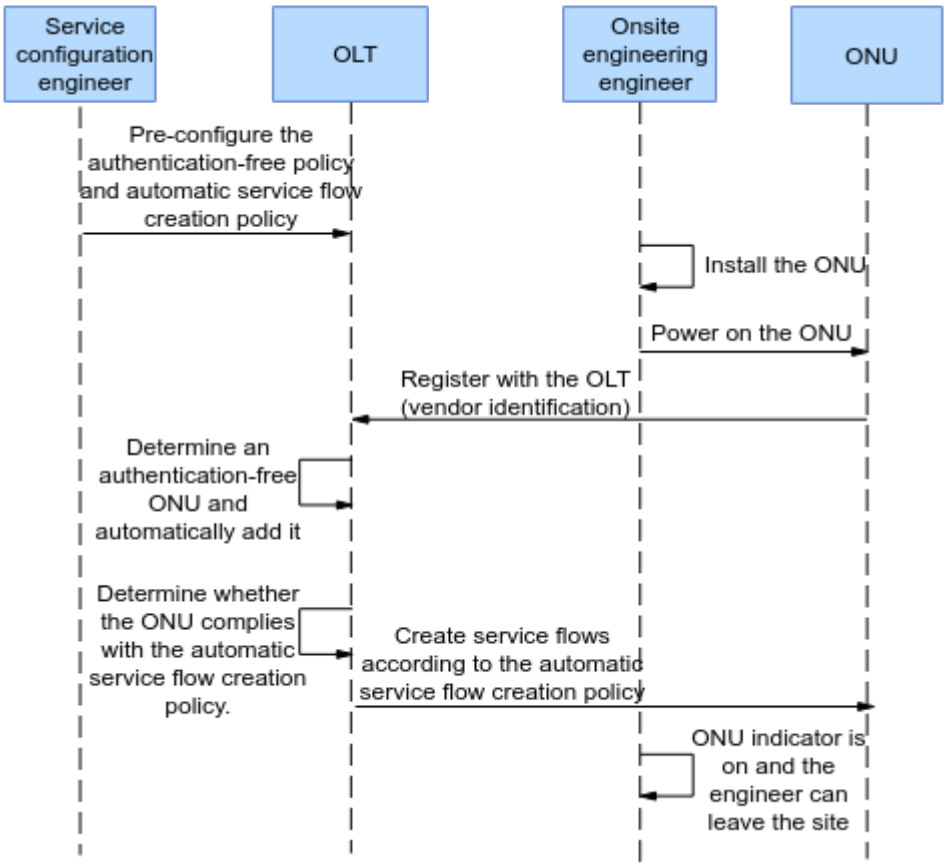
Automatic Service Flow Creation

The OLT supports pre-configuration of service flow creation policies on a PON port. Then, after an ONU goes online, services flows are automatically created based on the pre-configured policies. This function simplifies the configuration process and improves the installation and deployment efficiency.

O&M for these automatically created service flows is the same as that for commonly created service flows.

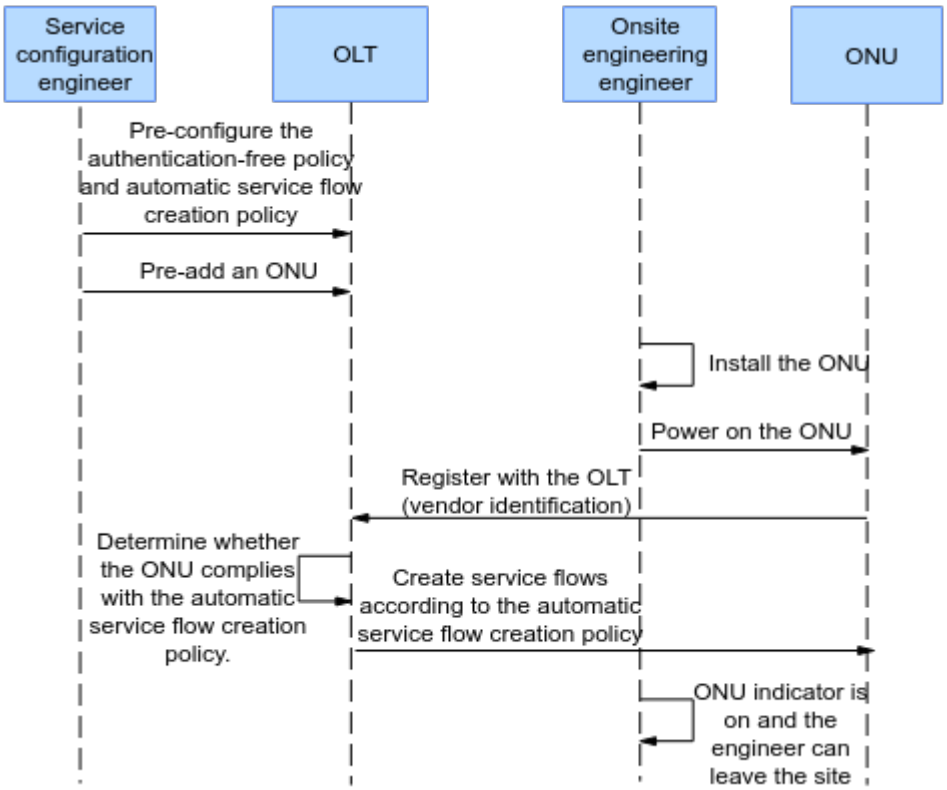
 Copyright © Huawei Technologies Co., Ltd.

Figure 10-20 Process of automatic service flow creation (no ONU is manually added)



 Copyright © Huawei Technologies Co., Ltd.

Figure 10-21 Process of automatic service flow creation (ONU is manually added)



10.3.3 Configuring Service Flow

The method of configuring service flows varies according to the access mode. You can locate the desired configuration description according to the access mode.

10.3.3.1 Creating an Ethernet Service Flow

An Ethernet service flow is a service channel connecting the user side to the network side. Configure a service flow before provisioning services.

Context

A service flow can carry a single service or multiple services. A multi-service flow is used to carry triple-play services. A service port can carry a single service flow or multiple service flows. If a service port carries multiple service flows, the OLT supports traffic classification and identifies users or services based on:

- User-side VLAN.
- User-side service encapsulation type. For example, PPPoE is used for the Internet access service and IPoE is used for multicast services.
- VLAN+user-side packet priority. For example, the priority of multicast services is higher than that of the Internet access service.
- VLAN+user-side service encapsulation type.

Before creating a service flow, run the **display traffic table** command to query whether the desired traffic profile has existed in the system. If no desired traffic

 Copyright © Huawei Technologies Co., Ltd.

profile is available, run the **traffic table ip** command to create a traffic profile for the service flow. The OLT provides seven default traffic profiles with IDs ranging from 0 to 6.

Table 10-21 lists the default settings of a service flow.

Table 10-21 Default settings of a service flow

Parameter	Default Setting
Traffic profile ID	0-6
Management status	Activated

Procedure

Step 1 Create a service flow or multiple service flows in batches.

- Run the **service-port** command to create a service flow.
  - Single-service flow:

    Select **single-service** or do not set **multi-service**.
  - Multi-service flow based on the user-side VLAN:

    Select **multi-service user-vlan** { **untagged** | *user-vlanid* | **priority-tagged** | **other-all** }.

    - **untagged**: If this parameter is specified, user packets do not carry a tag.
    - *user-vlanid*: If this parameter is specified, user packets carry a tag, which is the customer VLAN (C-VLAN).
    - **priority-tagged**: When **priority-tagged** is selected, the VLAN tag is 0 and the priorities of user-side packets are 0-7. (The highest priority is 7.)
    - **other-all**: If this parameter is specified, the created service flow carries QinQ transparent LAN service (TLS) services for enterprises. User packets are matched based on the specified user VLAN (or untagged attribute). The unmatched packets are transmitted on the TLS service flow to the upper-layer network.
  - Multi-service flow based on the user-side service encapsulation mode:

    Select **multi-service user-encap** *user-encap*.
  - Multi-service flow based on the VLAN+user-side packet priority (802.1p)

    Select **multi-service user-8021p** *user-8021p* [**user-vlan** *user-vlanid*].
  - Multi-service flow based on the VLAN+user-side service encapsulation mode (user-encap)

    Select **multi-service user-vlan** { **untagged** | *user-vlanid* | **priority-tagged** } **user-encap** *user-encap*.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

- The OLT supports the service flow configuration by index. Each service flow has a unique index. In this manner, users do not need to input a large number of flow parameters, thereby simplifying service flow configurations. When creating a service flow, the service flow index parameter *index* is optional. If *index* is not set, the OLT automatically allocates an idle index starting from the configured maximum index, regardless of whether the maximum index has been deleted. If the idle index is greater than the upper index threshold, the OLT searches for the new idle index starting from 0.
- **vlan** is an S-VLAN, which can only be a smart or MUX VLAN.
- The access mode can be ATM or PTM. In ATM access mode, the VPI and VCI must be input and the same as the VPI and VCI of the access terminal.
- **rx-cttr** is the same as **outbound** in terms of meaning and function. Either parameter indicates the index of the service flow from the network side to the user side. **tx-cttr** is the same as **inbound** in terms of meaning and function. Either parameter indicates the index of the service flow from the user side to the network side.

- Run the **multi-service-port** command to create service flows in batches.

Step 2 (Optional) Run the **service-port desc** command to configure the description of the service flow. The description includes the purpose of the service flow creation and the services carried over the service flow. Therefore, the description of the service flow facilitates maintenance.

Step 3 (Optional) Run the **service-port index adminstatus** command to configure the management status of the service flow. A service port activated by default.

A service can be provisioned at two levels: port level and service flow level. The service takes effect only after both the access port and service flow are activated.

----End

Example

The following configurations are used as an example to provision the Internet access service: The access port is 0/4/1. The ID of the user side VLAN is 2. The ID of the upstream VLAN is 100. The Internet access service uses traffic profile 8. The profile name is net, the CIR is set to 10240, the upstream priority is set to 0, and the packet priority selection policy is the traffic profile for scheduling based on the local priority.

```
huawei(config)#traffic table ip index 8 name net cir 10240 priority 0 priority-policy
 local-setting
  Create traffic descriptor record successfully
  ------------------------------------------------
  TD Index            : 8
  TD Name             : net
  Priority            : 0
  Copy Priority       : -
  CTAG Mapping Priority: -
  CTAG Default Priority: 0
  Priority Policy     : local-pri
  CIR                 : 10240 kbps
  CBS                 : 329680 bytes
  PIR                 : 20480 kbps
  PBS                 : 657360 bytes
  Color Mode          : color-blind
  Referenced Status   : not used
  ------------------------------------------------
huawei(config)#service-port 1 vlan 100 eth 0/4/1 multi-service user-vlan 2 rx-cttr 8 tx-cttr 8
```

 Copyright © Huawei Technologies Co., Ltd.

10.3.3.2 Creating a GPON Service Flow

A GPON service flow is a service channel connecting the user side to the network side. Configure a service flow before provisioning services.

Context

A service flow can carry a single service or multiple services. A multi-service flow is used to carry triple-play services. A service port can carry a single service flow or multiple service flows. If a service port carries multiple service flows, the OLT supports traffic classification and identifies users or services based on:

- User-side VLAN.
- User-side service encapsulation type.
- VLAN+user-side packet priority.
- VLAN+user-side service encapsulation type.

Table 10-22 lists the default settings of a service flow.

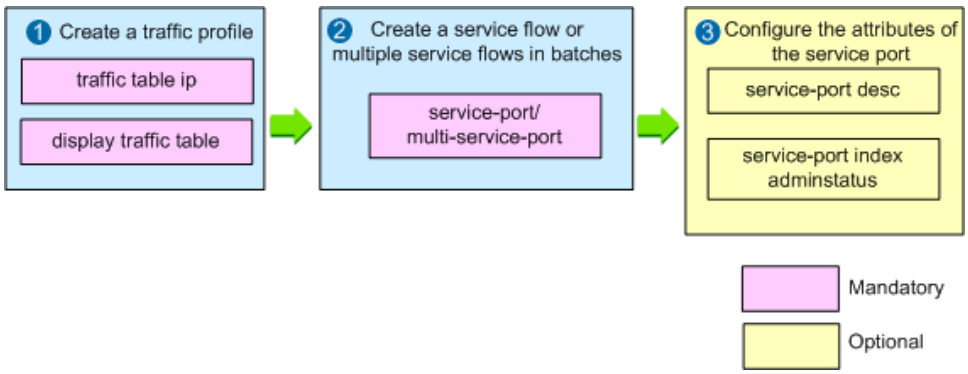
Table 10-22 Default settings of a service flow

Parameter	Default Setting
Traffic profile ID	0-6
Management status	Activated

Configuration Process

Figure 10-22 shows the process of creating a GPON service flow.

Figure 10-22 Process of creating a GPON service flow



Procedure

Step 1 Create a traffic profile.

Run the **traffic table ip** command to create a traffic profile. The OLT provides seven default traffic profiles with IDs ranging from 0 to 6.

Before creating a service flow, run the **display traffic table** command to query whether the desired traffic profile has existed in the system. If no desired traffic

 Copyright © Huawei Technologies Co., Ltd.

profile is available, run the **traffic table ip** command to create a traffic profile for the service flow.

Step 2 Create a service flow or multiple service flows in batches.

- Run the **service-port** command to create a service flow.
  - Single-service flow:

    Do not set **multi-service**.
  - Multi-service flow based on the user-side VLAN:

    Select **multi-service user-vlan** { **untagged** | *user-vlanid* | **priority-tagged** | **other-all** }.

    - **untagged**: If this parameter is specified, user packets do not carry a tag.
    - *user-vlanid*: If this parameter is specified, user packets carry a tag, which is the customer VLAN (C-VLAN).
    - **priority-tagged**: When priority-tagged is selected, the VLAN tag is 0 and the priorities of user-side packets are 0-7.
    - **other-all**: If this parameter is specified, the created service flow carries QinQ transparent LAN service (TLS) services for enterprises. User packets are matched based on the specified user VLAN (or untagged attribute). The unmatched packets are transmitted on the TLS service flow to the upper-layer network.
  - Multi-service flow based on the user-side service encapsulation mode:

    Select **multi-service user-encap** *user-encap*.
  - Multi-service flow based on the VLAN+user-side packet priority (802.1p)

    Select **multi-service user-8021p** *user-8021p* [**user-vlan** *user-vlanid*].
  - Multi-service flow based on the VLAN+user-side service encapsulation mode (user-encap)

    Select **multi-service user-vlan** { **untagged** | *user-vlanid* | **priority-tagged** } **user-encap** *user-encap*.

  NOTE

  - The OLT supports the service flow configuration by index. Each service flow has a unique index. In this manner, users do not need to input a large number of flow parameters, thereby simplifying service flow configurations. When creating a service flow, the service flow index parameter *index* is optional. If *index* is not set, the OLT automatically allocates an idle index starting from the minimum unused value.
  - **vlan** is an S-VLAN, which can only be a smart or MUX VLAN.
  - **rx-cttr** is the same as **outbound** in terms of meaning and function. Either parameter indicates the index of the service flow from the network side to the user side. **tx-cttr** is the same as **inbound** in terms of meaning and function. Either parameter indicates the index of the service flow from the user side to the network side. The traffic profile bound to the service flow is the one created in **Step 1**.
- Run the **multi-service-port** command to create service flows in batches.

Step 3 Configure the attributes of the service port. Configure the attributes of the service port according to requirements.

- Run the **service-port desc** command to configure the description of the service flow. The description includes the purpose of the service flow creation

 Copyright © Huawei Technologies Co., Ltd.

and the services carried over the service flow. Therefore, the description of the service flow facilitates maintenance.

- Run the **service-port index adminstatus** command to configure the management status of the service flow. A service port activated by default.

  A service can be provisioned at two levels: port level and service flow level. The service takes effect only after both the access port and service flow are activated.

----End

Example

The following configurations are used as an example to provision the Internet access service for a user:

- GPON port 0/2/0 on an OLT connects to an ONT.
- The planned configurations of the user are as follows:
  - The rate of the Internet access service is 4096 Kbit/s.
  - The index of the GEM port carrying the Internet access service is 135.
  - The ID of the service VLAN (S-VLAN) is 1000.
- No proper traffic profile is available and traffic profile 10 is created.
- The user has not registered. Do not provision the Internet access service for the user.

```
huawei(config)#traffic table ip index 10 cir 4096 priority 3 priority-policy local-Setting
huawei(config)#service-port 5 vlan 1000 gpon 0/2/0 ont 1 gemport 135 inbound traffic-table index 10 outbound traffic-table index 10
huawei(config)#service-port 5 adminstatus disable
```

The following configurations are used as an example to provision the Internet access service for a commercial user:

- GPON port 0/2/0 on an OLT connects to an ONT.
- The planned configurations of the user are as follows:
  - The rate of the Internet access service is 8192 Kbit/s.
  - The Internet access service is in multi-service mode to facilitate follow-up service expansion.
  - The user is identified by C-VLAN.
  - The ID of the S-VLAN is 1023.
  - The ID of the C-VLAN is 100.
  - The index of the GEM port carrying the Internet access service is 130.
- No proper traffic profile is available and traffic profile 8 is created.
- The description of the service flow is added to facilitate maintenance.
- The Internet access service needs to be provisioned immediately.

```
huawei(config)#traffic table ip index 8 cir 8192 priority 4 priority-policy local-Setting
huawei(config)#service-port 10 vlan 1023 gpon 0/2/0 gemport ont 1 130 multi-service user-vlan 100 inbound traffic-table index 8 outbound traffic-table index 8
huawei(config)#service-port desc 10 description gpon/vlanid:1023/uservlan:100
```

 Copyright © Huawei Technologies Co., Ltd.

10.3.3.3 Creating a GPON Service Flow (in Profile Mode with Universal Configurations)

A GPON service flow is a service channel connecting the user side to the network side. Configure a service flow before provisioning services.

Context

A service flow can carry a single service or multiple services. A multi-service flow is used to carry triple-play services. A service port can carry a single service flow or multiple service flows. If a service port carries multiple service flows, the OLT supports traffic classification and identifies users or services based on:

- User-side VLAN.
- User-side service encapsulation type.
- VLAN+user-side packet priority.
- VLAN+user-side service encapsulation type.

Table 10-23 lists the default settings of a service flow.

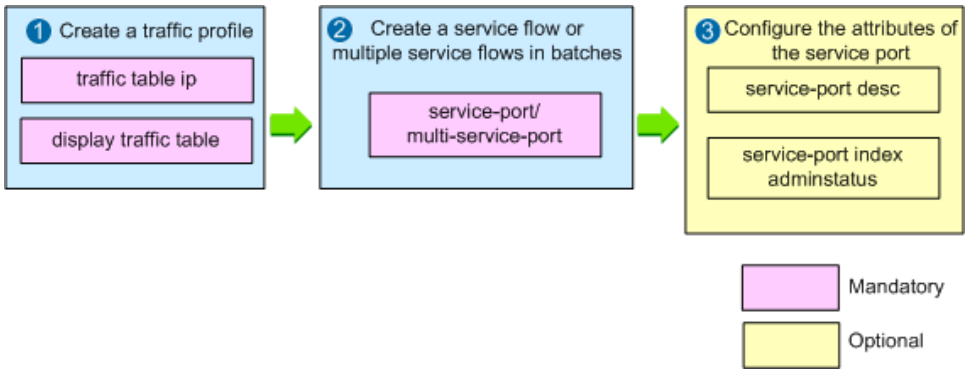
Table 10-23 Default settings of a service flow

Parameter	Default Setting
Traffic profile ID	0-6
Management status	Activated

Configuration Process

Figure 10-23 shows the process of creating a GPON service flow.

Figure 10-23 Process of creating a GPON service flow



Procedure

Step 1 Create a traffic profile.

Run the **traffic table ip** command to create a traffic profile. The OLT provides seven default traffic profiles with IDs ranging from 0 to 6.

 Copyright © Huawei Technologies Co., Ltd.

Before creating a service flow, run the **display traffic table** command to query whether the desired traffic profile has existed in the system. If no desired traffic profile is available, run the **traffic table ip** command to create a traffic profile for the service flow.

Step 2 Create a service flow or multiple service flows in batches.

- Run the **service-port** command to create a service flow.
  - Single-service flow:

    Do not set **multi-service**.
  - Multi-service flow based on the user-side VLAN:

    Select **multi-service user-vlan** { **untagged** | *user-vlanid* | **priority-tagged** | **other-all** }.

    - **untagged**: If this parameter is specified, user packets do not carry a tag.
    - *user-vlanid*: If this parameter is specified, user packets carry a tag, which is the customer VLAN (C-VLAN).
    - **priority-tagged**: When priority-tagged is selected, the VLAN tag is 0 and the priorities of user-side packets are 0-7.
    - **other-all**: If this parameter is specified, the created service flow carries QinQ transparent LAN service (TLS) services for enterprises. User packets are matched based on the specified user VLAN (or untagged attribute). The unmatched packets are transmitted on the TLS service flow to the upper-layer network.
  - Multi-service flow based on the user-side service encapsulation mode:

    Select **multi-service user-encap** *user-encap*.
  - Multi-service flow based on the VLAN+user-side packet priority (802.1p)

    Select **multi-service user-8021p** *user-8021p* [**user-vlan** *user-vlanid*].
  - Multi-service flow based on the VLAN+user-side service encapsulation mode (user-encap)

    Select **multi-service user-vlan** { **untagged** | *user-vlanid* | **priority-tagged** } **user-encap** *user-encap*.

  NOTE

  - The OLT supports the service flow configuration by index. Each service flow has a unique index. In this manner, users do not need to input a large number of flow parameters, thereby simplifying service flow configurations. When creating a service flow, the service flow index parameter *index* is optional. If *index* is not set, the OLT automatically allocates an idle index starting from the configured maximum index, regardless of whether the maximum index has been deleted. If the idle index is greater than the upper index threshold, the OLT searches for the new idle index starting from 0.
  - **vlan** is an S-VLAN, which can only be a smart or MUX VLAN.
  - **rx-cttr** is the same as **outbound** in terms of meaning and function. Either parameter indicates the index of the service flow from the network side to the user side. **tx-cttr** is the same as **inbound** in terms of meaning and function. Either parameter indicates the index of the service flow from the user side to the network side. The traffic profile bound to the service flow is the one created in **Step 1**.
- Run the **multi-service-port** command to create service flows in batches.

 Copyright © Huawei Technologies Co., Ltd.

Step 3 Configure the attributes of the service port. Configure the attributes of the service port according to requirements.

- Run the **service-port desc** command to configure the description of the service flow. The description includes the purpose of the service flow creation and the services carried over the service flow. Therefore, the description of the service flow facilitates maintenance.
- Run the **service-port index adminstatus** command to configure the management status of the service flow. A service port activated by default.

  A service can be provisioned at two levels: port level and service flow level. The service takes effect only after both the access port and service flow are activated.

----End

Example

The following configurations are used as an example to provision the Internet access service for a user:

- GPON port 0/2/0 on an OLT connects to ONT 1.
- The planned configurations of the user are as follows:
  - The rate of the Internet access service is 4096 Kbit/s.
  - The index of the GEM port carrying the Internet access service is 126.
  - The ID of the service VLAN (S-VLAN) is 1000.
- No proper traffic profile is available and traffic profile 10 is created.
- The user has not registered. Do not provision the Internet access service for the user.

```
huawei(config)#traffic table ip index 10 cir 4096 priority 3 priority-policy loc
al-Setting
  Create traffic descriptor record successfully
  ------------------------------------------------
  TD Index            : 10
  TD Name             : ip-traffic-table_10
  Priority            : 3
  Mapping Priority    : -
  Mapping Index       : -
  CTAG Mapping Priority: -
  CTAG Mapping Index  : -
  CTAG Default Priority: 0
  Priority Policy     : local-pri
  CIR                 : 4096 kbps
  CBS                 : 133072 bytes
  PIR                 : 8192 kbps
  PBS                 : 264144 bytes
  Fix                 : 0 kbps
  CAR Threshold Profile: -
  Color Mode          : color-blind
  Color policy        : dei
  Referenced Status   : not used
  ------------------------------------------------
huawei(config)#service-port 5 vlan 1000 gpon 0/2/0 ont 1 gemport 126 inbound
traffic-table index 10 outbound traffic-table index 10
huawei(config)#service-port 5 adminstatus disable
```

The following configurations are used as an example to provision the Internet access service for a commercial user:

 Copyright © Huawei Technologies Co., Ltd.

- GPON port 0/2/0 on an OLT connects to ONT 2.
- The planned configurations of the user are as follows:
  - The rate of the Internet access service is 8192 Kbit/s.
  - The Internet access service is in multi-service mode to facilitate follow-up service expansion.
  - The user is identified by C-VLAN.
  - The ID of the S-VLAN is 1023.
  - The ID of the C-VLAN is 100.
  - The index of the GEM port carrying the Internet access service is 126.
- No proper traffic profile is available and traffic profile 8 is created.
- The description of the service flow is added to facilitate maintenance.
- The Internet access service needs to be provisioned immediately.

```
huawei(config)#display traffic table ip from-index 0
{ <cr>|to-index<K> }:

 Command:
        display traffic table ip from-index 0
 ------------------------------------------------------------------------------
  TID CIR      CBS      PIR      PBS      Pri Copy-policy         Pri-Policy
      (kbps)   (bytes)  (kbps)   (bytes)
 ------------------------------------------------------------------------------
    0 1024     34768    2048     69536      6 -                    tag-pri
    1 2496     81872    4992     163744     6 -                    tag-pri
    2 512      18384    1024     36768      0 -                    tag-pri
    3 576      20432    1152     40864      2 -                    tag-pri
    4 64       4048     128      8096       4 -                    tag-pri
    5 2048     67536    4096     135072     0 -                    tag-pri
    6 off      off      off      off        0 -                    tag-pri
 ------------------------------------------------------------------------------
 Total Num : 7
huawei(config)#traffic table ip index 8 cir 8192 priority 4 priority-policy
 local-Setting
 Create traffic descriptor record successfully
 ------------------------------------------------
 TD Index             : 8
 TD Name              : ip-traffic-table_8
 Priority             : 4
 Copy Priority        : -
 Mapping Index        : -
 CTAG Mapping Priority: -
 CTAG Mapping Index   : -
 CTAG Default Priority: 0
 Priority Policy      : local-pri
 CIR                  : 8192 kbps
 CBS                  : 264144 bytes
 PIR                  : 16384 kbps
 PBS                  : 526288 bytes
 Fix                  : 0 kbps
 CAR Threshold Profile: -
 Color Mode           : color-blind
 Color policy         : dei
 Referenced Status    : not used
 ------------------------------------------------
huawei(config)#service-port 10 vlan 1023 gpon 0/2/0 ont 2 gemport 126 multi-service
 user-vlan 100 inbound traffic-table index 8 outbound traffic-table index 8
huawei(config)#service-port desc 10 description gpon/Vlanid:1023/uservlan:100
```

 Copyright © Huawei Technologies Co., Ltd.

10.3.3.4 Creating a GPON Service Flow in Profile Mode with Simplified Configurations

A GPON service flow is a service channel connecting the user side to the network side. Configure a service flow before provisioning services. In a simplified service flow, the system only cares about only the translation between the SVLAN and the CVLAN and does not care about the GEM port mapping.

Context

A service port can carry a single service or multiple services. When a service port carries multiple services, the OLT supports traffic classification. In simplified configuration mode, the OLT supports only C-VLAN-based traffic classification.

Table 10-24 lists the default settings of a service flow.

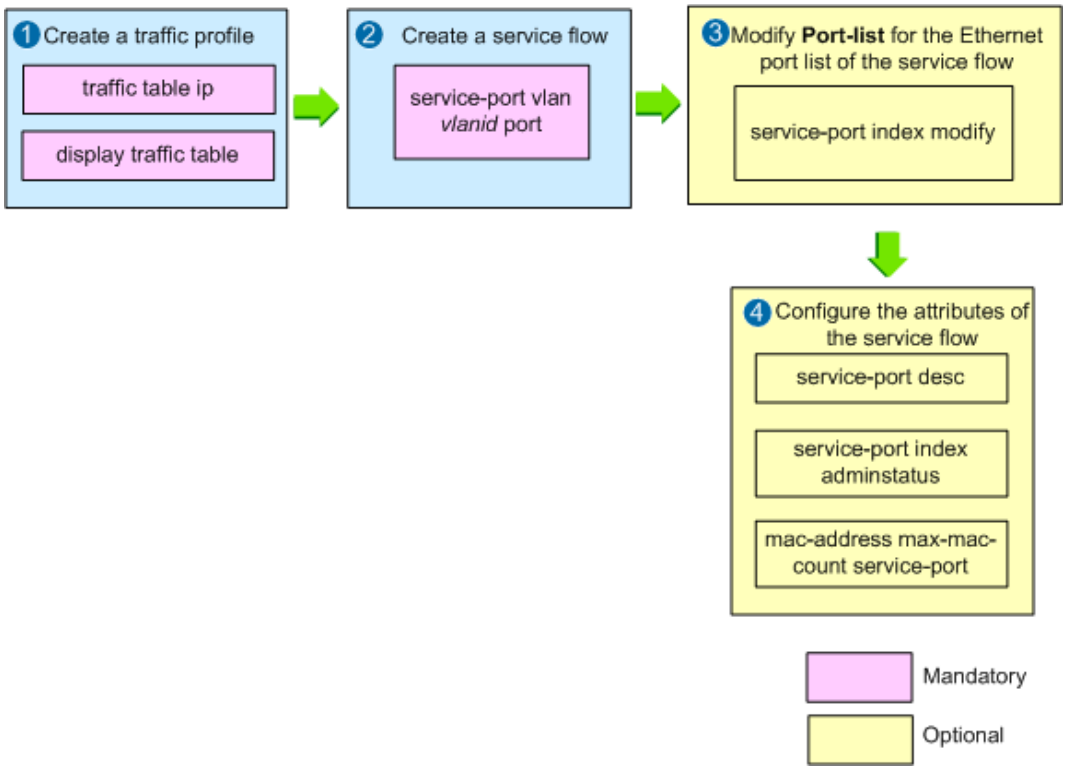
Table 10-24 Default settings of a service flow

Parameter	Default Setting
Traffic profile ID	0-6
Management status	Activated
Maximum number of learnable MAC addresses	1023

Configuration Process

Figure 10-24 shows the process of creating a GPON service flow.

Figure 10-24 Process of creating a GPON service flow



 Copyright © Huawei Technologies Co., Ltd.

Procedure

Step 1 Run the **traffic table ip** command to create a traffic profile.

The OLT provides seven default traffic profiles with IDs ranging from 0 to 6.

Before creating a service flow, run the **display traffic table** command to check whether a traffic profile meeting service requirement is available. If no such a traffic profile is available, create a traffic profile.

Step 2 Run the **service-port vlan vlanid port** command to create a service flow.

In simplified configuration mode, a service flow is an end-to-end configuration to a port on an optical network terminal (ONT) from the OLT. Then, users only need to pay attention to the switching between the S-VLAN and C-VLAN.

1. Select the type of an ONT port from **ont** *ontid* { **eth** | **iphost** }.
   - **eth**: an Ethernet port, which is used for the Internet access or multicast service.
   - **iphost**: a POTS port, which is used for the voice service.
2. Select a C-VLAN tag from **multi-service user-vlan** { **untagged** | *user-vlanid* }.
   - **untagged**: indicates that the user-side packets are untagged.
   - *user-vlanid*: indicates that the user-side packets are tagged. The tag value is the same as the C-VLAN tag.

The preceding parameters are set based on the ONT type. Therefore, specify the ONT type before creating a service flow.

- For a bridging ONT (SFU), the ONT port provisioning the POTS service is set to **iphost** and ONT ports provisioning other services are set to **eth**.
- For a gateway ONT (HGU), the ONT port provisioning the Internet access service with untagged packets is set to **iphost**; other services cannot be configured on physical ports but on the ONT.

Table 10-25 lists the commands involved in parameter settings.

Table 10-25 Commands involved in parameter settings

Service Type	Bridging ONT	Gateway ONT
Tagged high-speed Internet (HSI) service	**service-port** [*index*] **vlan** *vlanid* **port** *frameid/slotid/portid* **ont** *ontid* **eth** *port-index-list* **multi-service**...	**service-port** [*index*] **vlan** *vlanid* **port** *frameid/slotid/portid* **ont** *ontid* **multi-service**...
Untagged HSI service	**service-port** [*index*] **vlan** *vlanid* **port** *frameid/slotid/portid* **ont** *ontid* **eth** *port-index-list* **multi-service**...	**service-port** [*index*] **vlan** *vlanid* **port** *frameid/slotid/portid* **ont** *ontid* **iphost multi-service**... **NOTE** Services can be identified only after a native VLAN is configured for ONT ports.

 Copyright © Huawei Technologies Co., Ltd.

Service Type	Bridging ONT	Gateway ONT
Voice service	**service-port** [*index*] **vlan** *vlanid* **port** *frameid/slotid/portid* **ont** *ontid* **iphost multi-service**...	**service-port** [*index*] **vlan** *vlanid* **port** *frameid/slotid/portid* **ont** *ontid* **multi-service**...
IPTV service	**service-port** [*index*] **vlan** *vlanid* **port** *frameid/slotid/portid* **ont** *ontid* **eth** *port-index-list* **multi-service**...	**service-port** [*index*] **vlan** *vlanid* **port** *frameid/slotid/portid* **ont** *ontid* **multi-service**...

In the preceding table, "..." indicates the omitted command format.

NOTE

- The OLT supports the service flow configuration by index. Each service flow has a unique index. In this manner, users do not need to input a large number of flow parameters, thereby simplifying service flow configurations. When creating a service flow, the service flow index parameter *index* is optional. If *index* is not set, the OLT automatically allocates an idle index starting from the configured maximum index, regardless of whether the maximum index has been deleted. If the idle index is greater than the upper index threshold, the OLT searches for the new idle index starting from 0.
- **vlan** is an S-VLAN, which can only be a smart or MUX VLAN.
- **rx-cttr** is the same as **outbound** in terms of meaning and function. Either parameter indicates the index of the service flow from the network side to the user side. **tx-cttr** is the same as **inbound** in terms of meaning and function. Either parameter indicates the index of the service flow from the user side to the network side. The traffic profile bound to the service flow is the one created in **Step 1**.

Step 3 (Optional) Run the **service-port index modify** command to modify **Port-list** for the Ethernet port list of the service flow.

Perform this step if the service planning for a user port is changed.

Step 4 (Optional) Configure the attributes of the service flow.

- Run the **service-port desc** command to configure the description of the service flow. The description includes the purpose of the service flow creation and the services carried over the service flow. Therefore, the description of the service flow facilitates maintenance.
- Run the **service-port index adminstatus** command to configure the management status of the service flow. A service port activated by default.

  A service can be provisioned at two levels: port level and service flow level. The service takes effect only after both the access port and service flow are activated.
- Run the **mac-address max-mac-count service-port** command to set the maximum number of MAC addresses learned by the service port. This configuration restricts the maximum number of PCs that can access the Internet by using the same user account. The maximum number of learnable MAC addresses of a service flow is 1023 by default.

----End

 Copyright © Huawei Technologies Co., Ltd.

Example

The following configurations are used as an example to provision the Internet access service for a user:

- GPON port 0/2/0 on an OLT connects to ONT 1.
- The planned configurations of the user are as follows:
  - The rate of the Internet access service is 4096 kbit/s.
  - The user connects to Ethernet port 1 on the ONT.
  - The ID of the service VLAN (S-VLAN) is 1000.
  - At most three users can concurrently access the Internet using the same account.
- No proper traffic profile is available and traffic profile **hsi** is created.
- The user has not registered. Do not provision the Internet access service for the user.

```
huawei(config)#traffic table ip name hsi cir 4096 priority 3 priority-policy loc
al-Setting
  Create traffic descriptor record successfully
  ------------------------------------------------
  TD Index            : 10
  TD Name             : hsi
  Priority            : 3
  Copy Priority       : -
  Mapping Index       : -
  CTAG Mapping Priority: -
  CTAG Mapping Index  : -
  CTAG Default Priority: 0
  Priority Policy     : local-pri
  CIR                 : 4096 kbps
  CBS                 : 133072 bytes
  PIR                 : 8192 kbps
  PBS                 : 264144 bytes
  Fix                 : 0 kbps
  CAR Threshold Profile: -
  Color Mode          : color-blind
  Color policy        : dei
  Referenced Status   : not used
  ------------------------------------------------
huawei(config)#service-port 5 vlan 1000 gpon 0/2/0 ont 1 gemport 120 inbound traffic-table name hsi
outbound traffic-table name hsi
huawei(config)#mac-address max-mac-count service-port 5 3
huawei(config)#service-port 5 adminstatus disable
```

The following configurations are used as an example to provision the Internet access service for a commercial user:

- GPON port 0/2/0 on an OLT connects to ONT 2.
- The planned configurations of the user are as follows:
  - The rate of the Internet access service is 8192 kbit/s.
  - The Internet access service is in multi-service mode to facilitate follow-up service expansion.
  - The user is identified by customer VLAN (C-VLAN).
  - The ID of the S-VLAN is 1023.
  - The ID of the C-VLAN is 100.
  - ONT port 2 connects to the service flow.
- No proper traffic profile is available and traffic profile **internet** is created.

 Copyright © Huawei Technologies Co., Ltd.

- The description of the service flow is added to facilitate maintenance.
- The Internet access service needs to be provisioned immediately.

```
huawei(config)#display traffic table ip from-index 0
{ <cr>|to-index<K> }:

 Command:
        display traffic table ip from-index 0
 ------------------------------------------------------------------------------
  TID CIR(kbps) CBS(bytes) PIR(kbps) PBS(bytes) Pri Copy-policy    Pri-Policy
 ------------------------------------------------------------------------------
    0     1024      34768      2048      69536   6 -             tag-pri
    1     2496      81872      4992     163744   6 -             tag-pri
    2      512      18384      1024      36768   0 -            tag-pri
    3      576      20432      1152      40864   2 -            tag-pri
    4       64       4048       128       8096   4 -           tag-pri
    5     2048      67536      4096     135072   0 -             tag-pri
    6      off        off       off        off   0 -           tag-pri
 ------------------------------------------------------------------------------
  Total Num : 7
huawei(config)#traffic table ip name internet cir 8192 priority 4 priority-policy local-Setting
  Create traffic descriptor record successfully
  ------------------------------------------------
  TD Index             : 8
  TD Name              : internet
  Priority             : 4
  Copy Priority        : -
  Mapping Index        : -
  CTAG Mapping Priority: -
  CTAG Mapping Index   : -
  CTAG Default Priority: 0
  Priority Policy      : local-pri
  CIR                  : 8192 kbps
  CBS                  : 264144 bytes
  PIR                  : 16384 kbps
  PBS                  : 526288 bytes
  Fix                  : 0 kbps
  CAR Threshold Profile: -
  Color Mode           : color-blind
  Color policy         : dei
  Referenced Status    : not used
  ------------------------------------------------
huawei(config)#service-port 10 vlan 1023 port 0/2/0 ont 2 eth 2 multi-service user-vlan 100 inbound
traffic-table name internet outbound traffic-table name internet
huawei(config)#service-port desc 10 description gpon/Vlanid:1023/uservlan:100
```

A home user initially configures the Internet access service on ONT port 1. Later, the user needs to use the Internet access service ONT ports 2 and 3. The original service flow ID is 10.

```
huawei(config)#service-port 10 modify ont eth 1-3
```

10.3.3.5 Configuring Automatic Service Flow Creation (GPON)

The OLT supports configuration of service flow creation policies on PON ports. After an ONU goes online under the PON port, service flows are automatically created according to the preset policies. This topic uses MA5671-24FE as an example for the ONU. Configuration methods are similar for other device models.

 Copyright © Huawei Technologies Co., Ltd.

Data Plan

Table 10-26 Key data plan

Item	Data
Networking data	Upstream port: 0/9/0 PON port: 0/1/0 ONU ID: 1
VLAN	Internet access VLAN (SVLAN): ● SVLAN ID: 100 ● SVLAN type: smart ● For the SVLAN planned with single-tagged VLAN, the attribute of the SVLAN is common by default and no configuration is required. For the SVLAN planned with double-tagged VLNA, the attribute of the SVLAN is QinQ or stacking. For a PC connected on the user side, user packets are untagged.
Traffic profile	Profile name: ftto_hsi CIR: 10 Mbit/s PIR: 20 Mbit/s Priority policy: user-cos Default 802.1p priority of packets: 0 Enqueuing priority scheduling policy of packets: Tag-In-Package
Automatic flow creation policy	Maximum number of flows automatically created on the ONU: 24 (If the actual flow number exceeds 24, extra flows can still be created successfully. However, an alarm, which indicates that the flow number range is exceeded, is reported.) Upstream traffic profile: ftto_hsi Downstream traffic profile: ftto_hsi VLAN translation policy: ● When a service flow with the single-tagged VLAN is created, all ONUs have the same CVLAN range, the OLT implements VLAN translation, and the SVLAN is added in the upstream direction. That is, CVLAN<->SVLAN. ● When a service flow with the double-tagged VLAN is created, all ONUs have the same CVLAN range, and the OLT implements VLAN translation. The start value of inner VLAN (C'VLAN) on the network side is 2000. The SVLAN and C'VLAN are added in the upstream direction. That is, CVLAN<->SVLAN + C'VLAN.
DBA profile	Profile name: ftto_dba Profile type: Type3 Assured bandwidth: 10 Mbit/s Maximum bandwidth: 20 Mbit/s

 Copyright © Huawei Technologies Co., Ltd.

Item	Data
ONU line profile	Profile name: ftto_line T-CONT ID: 1 GEM Port ID for the Internet access service: 1, 2, or 3 **NOTE** For example, to deploy services on 24 Ethernet ports, 3 GEM ports need to be configured, because each GEM port can be mapped to 8 Ethernet ports.
ONU service profile	Profile name: ftto_ser ONU port capability set: 24
ONU automatic adding policy	Device type: MA5671-F24 ONU line profile name: ftto_line ONU service profile name: ftto_ser

Procedure

Step 1 Configure an Internet access service VLAN (SVLAN).

Create Internet service VLAN 100, and add upstream port 0/9/0 to VLAN 100.

```
huawei(config)#vlan 100 smart
huawei(config)#port vlan 100 0/9 0
```

If the double-tagged SVLAN is planned, perform the following operation. If the single-tagged VLAN is planned, no configuration is required.

```
huawei(config)#vlan attrib 100 q-in-q
```

Step 2 Configure a traffic profile.

Run the **display traffic table ip** command to query existing traffic profiles in the system. If the existing traffic profiles in the system do not meet the requirements, run the **traffic table ip** command to add a traffic profile.

```
huawei(config)#traffic table ip name ftto_hsi cir 10240 pir 20480 priority user-cos 0 priority-policy tag-In-package
```

Step 3 Configure an automatic flow creation policy.

A PON port can be configured with one policy at one time. The policy is configured based on a PON port or board. In this example, configure the automatic flow creation policy on GPON port 0/1/0.

- Create a service flow with the single-tagged VLAN.

  Create a service flow with the single-tagged VLAN on an ONU, add a specified SVLAN tag in the upstream direction, and remove the tag in the downstream direction.

  ```
  huawei(config)#interface gpon 0/1
  huawei(config-if-gpon-0/1)#auto-service-port port 0 vlan 100 single-vlan onu-vlan-num 24 inbound traffic-table name ftto_hsi outbound traffic-table name ftto_hsi
  huawei(config-if-gpon-0/1)#quit
  ```

- Create a service flow with the double-tagged VLAN.

 Copyright © Huawei Technologies Co., Ltd.

Create a service flow for each port of the ONU and add the C 'VLAN and SVLAN tags to the service flow in the upstream direction. The C' VLAN is calculated as Start VLAN + ONU ID x onu-vlan-num. The C' VLAN ID of the port is increased by 1. For example, if the start VLAN is 2000, ONU ID is 1, and onu-vlan-num is 24, C'VLAN ID = 2000 + 1 x 24 = 2024. For the 24 Ethernet ports of the ONU the value range of C'VLAN is 2024–2047.

```
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#auto-service-port port 0 vlan 100 double-vlan from-vlan 2000 onu-vlan-num 24 inbound traffic-table name ftto_hsi outbound traffic-table name ftto_hsi
huawei(config-if-gpon-0/1)#quit
```

Step 4 Configure GPON ONU profiles.

GPON ONU profiles include the DBA profile, line profile, service profile, and alarm profile.

- DBA profile: A DBA profile describes GPON traffic parameters. A T-CONT is bound to a DBA profile for dynamic bandwidth allocation, improving upstream bandwidth usage.
- Line profile: A line profile describes the binding between the T-CONT and the DBA profile, the QoS mode of the service flow, and the mapping between the GEM port and the ONU-side service.
- Service profile: A service profile provides the service configuration channel for the ONU that is managed through OMCI.
- Alarm profile: An alarm profile contains a series of alarm thresholds to measure and monitor the performance of activated ONU lines. When a statistical value reaches the threshold, the NE is notified and an alarm is reported to the log NE and the NMS.

1. Configure a DBA profile.

   Run the **display dba-profile** command to query the existing DBA profiles in the system. If the DBA profiles existing in the system cannot meet the requirements, you need to run the **dba-profile add** command to add one.

   ```
   huawei(config)#dba-profile add profile-name ftto_dba type3 assure 10240 max 20480
   ```

   NOTE

   Select a DBA profile of the proper bandwidth type and configure proper bandwidths according to the service types and total number of users of the ONU. Note that the sum of the fixed bandwidth and the assured bandwidth must not be greater than the total bandwidth of the PON port.

2. Configure an ONU line profile.

   Create a GPON ONU line profile named **ftto_line** and bind it to DBA profile **ftto_dba**.

   ```
   huawei(config)#ont-lineprofile gpon profile-name ftto_line
   huawei(config-gpon-lineprofile-1)#tcont 1 dba-profile-name ftto_dba
   ```

   Bind GEM ports 1, 2, and 3 to T-CONT 1.

   ```
   huawei(config-gpon-lineprofile-1)#gem add 1 eth tcont 1
   huawei(config-gpon-lineprofile-1)#gem add 2 eth tcont 1
   huawei(config-gpon-lineprofile-1)#gem add 3 eth tcont 1
   ```

   Configure the mapping between 3 GEM ports and 24 CVLANs.

 Copyright © Huawei Technologies Co., Ltd.

- When user packets are untagged and the MA5671 adopts the automatic adding policy, all Ethernet ports are configured with different native VLANs. The value of the native VLAN is the Ethernet port ID.
- When user packets carry CVLAN tags, configure relevant VLANs for mapping on the line profile. That is, replace VLAN IDs with corresponding CVLAN IDs. For example, when Ethernet port 1 on the MA5671 receives user packets with CVLAN 24, run the **gem mapping 1 0 vlan 24** command.

```
huawei(config-gpon-lineprofile-1)#gem mapping 1 0 vlan 1
huawei(config-gpon-lineprofile-1)#gem mapping 1 1 vlan 2
huawei(config-gpon-lineprofile-1)#gem mapping 1 2 vlan 3
huawei(config-gpon-lineprofile-1)#gem mapping 1 3 vlan 4
huawei(config-gpon-lineprofile-1)#gem mapping 1 4 vlan 5
huawei(config-gpon-lineprofile-1)#gem mapping 1 5 vlan 6
huawei(config-gpon-lineprofile-1)#gem mapping 1 6 vlan 7
huawei(config-gpon-lineprofile-1)#gem mapping 1 7 vlan 8
huawei(config-gpon-lineprofile-1)#gem mapping 2 0 vlan 9
huawei(config-gpon-lineprofile-1)#gem mapping 2 1 vlan 10
huawei(config-gpon-lineprofile-1)#gem mapping 2 2 vlan 11
huawei(config-gpon-lineprofile-1)#gem mapping 2 3 vlan 12
huawei(config-gpon-lineprofile-1)#gem mapping 2 4 vlan 13
huawei(config-gpon-lineprofile-1)#gem mapping 2 5 vlan 14
huawei(config-gpon-lineprofile-1)#gem mapping 2 6 vlan 15
huawei(config-gpon-lineprofile-1)#gem mapping 2 7 vlan 16
huawei(config-gpon-lineprofile-1)#gem mapping 3 0 vlan 17
huawei(config-gpon-lineprofile-1)#gem mapping 3 1 vlan 18
huawei(config-gpon-lineprofile-1)#gem mapping 3 2 vlan 19
huawei(config-gpon-lineprofile-1)#gem mapping 3 3 vlan 20
huawei(config-gpon-lineprofile-1)#gem mapping 3 4 vlan 21
huawei(config-gpon-lineprofile-1)#gem mapping 3 5 vlan 22
huawei(config-gpon-lineprofile-1)#gem mapping 3 6 vlan 23
huawei(config-gpon-lineprofile-1)#gem mapping 3 7 vlan 24
```

After the configurations are complete, run the **commit** command to apply the parameters setting.

```
huawei(config-gpon-lineprofile-1)#commit
huawei(config-gpon-lineprofile-1)#quit
```

3. Configure an ONU service profile.

   Create a GPON ONU service profile named **ftto_ser**. Configure the Ethernet port capability set to 24.

```
huawei(config)#ont-srvprofile gpon profile-name ftto_ser
huawei(config-gpon-srvprofile-1)#ont-port eth 24
```

   When user packets carry CVLAN tags, replace VLAN IDs with corresponding CVLAN IDs. For example, when Ethernet port 1 on the MA5671 receives user packets with CVLAN 24, run the **port vlan eth 1 24** command.

```
huawei(config-gpon-srvprofile-1)#port vlan eth 1 1
huawei(config-gpon-srvprofile-1)#port vlan eth 2 2
huawei(config-gpon-srvprofile-1)#port vlan eth 3 3
huawei(config-gpon-srvprofile-1)#port vlan eth 4 4
huawei(config-gpon-srvprofile-1)#port vlan eth 5 5
huawei(config-gpon-srvprofile-1)#port vlan eth 6 6
huawei(config-gpon-srvprofile-1)#port vlan eth 7 7
huawei(config-gpon-srvprofile-1)#port vlan eth 8 8
huawei(config-gpon-srvprofile-1)#port vlan eth 9 9
huawei(config-gpon-srvprofile-1)#port vlan eth 10 10
huawei(config-gpon-srvprofile-1)#port vlan eth 11 11
huawei(config-gpon-srvprofile-1)#port vlan eth 12 12
huawei(config-gpon-srvprofile-1)#port vlan eth 13 13
huawei(config-gpon-srvprofile-1)#port vlan eth 14 14
huawei(config-gpon-srvprofile-1)#port vlan eth 15 15
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-gpon-srvprofile-1)#port vlan eth 16 16
huawei(config-gpon-srvprofile-1)#port vlan eth 17 17
huawei(config-gpon-srvprofile-1)#port vlan eth 18 18
huawei(config-gpon-srvprofile-1)#port vlan eth 19 19
huawei(config-gpon-srvprofile-1)#port vlan eth 20 20
huawei(config-gpon-srvprofile-1)#port vlan eth 21 21
huawei(config-gpon-srvprofile-1)#port vlan eth 22 22
huawei(config-gpon-srvprofile-1)#port vlan eth 23 23
huawei(config-gpon-srvprofile-1)#port vlan eth 24 24
```

After the configurations are complete, run the **commit** command to apply the parameters setting.

```
huawei(config-gpon-srvprofile-1)#commit
huawei(config-gpon-srvprofile-1)#quit
```

4. (Optional) Add an alarm threshold profile.
   - The ID of the default GPON alarm threshold profile is 1. The thresholds of all the alarm parameters in the default alarm profile are 0, which indicates that no alarm is reported.
   - In this example, the default alarm threshold profile is used and therefore, the configuration of the alarm profile is not required.
   - To configure the alarm threshold parameters for monitoring the performance statistics of an activated ONU line, run the **gpon alarm-profile add** command to configure a GPON alarm threshold profile.

Step 5 Configure an ONT automatic adding policy.

The value of **equipment** is determined based on the actual device type.

Different types of ONTs must be configured with different automatic adding policies. The MA5671, which has 24 FE ports and supports local power supply, is used an example.

When **auto-confirm** is **enable**, do not need to run the **ont confirm** command on the GPON port to confirm the ONT. When **auto-confirm** is **disable**, run the **ont confirm** command on the GPON port.

```
huawei(config)#ont auto-add-policy gpon equipment MA5671-F24 omci ont-lineprofile-name ftto_line ont-srvprofile-name ftto_ser auto-confirm enable
```

After the configuration is complete, run the **display ont auto-add-policy** command to query the information.

```
huawei(config)#display ont auto-add-policy gpon all
------------------------------------------------------------------------------
 Index                : 1
 Ont EquipmentID      : MA5671-F24(0x4d41353637312d463234)
 PON mode             : GPON
 Management mode      : OMCI
 Line profile ID      : 1
 Line profile name    : ftto_line
 Service profile ID   : 1
 Service profile name : ftto_ser
 Auto-confirm         : enable
------------------------------------------------------------------------------
```

Step 6 (Optional) When the device supports reverse power over Ethernet (PoE), configure reversed PoE work modes for the ONU.

There are 3 reversed PoE work modes:

- **all-port**: Reversed PoE is provided for all ports regardless of the powered port.

 Copyright © Huawei Technologies Co., Ltd.

- **powered-port**: Reversed PoE is provided for the powered port only. Non-powered ports are shut down and services on them are interrupted.
- **unconcern**: The OLT does not issue reversed PoE configuration to the ONU. The default value is **unconcern**.

```
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#ont reverse-poe work-mode 0 0 all-port
huawei(config-if-gpon-0/1)#quit
```

Run the **display ont reverse-poe** command to query the reversed PoE work mode of the ONU and reversed PoE status on all Ethernet ports.

```
huawei(config-if-gpon-0/1)#display ont reverse-poe 0 0
  ---------------------------------------------------------------------------
  Reverse PoE work mode   : All ports
  ---------------------------------------------------------------------------
  Port-ID    Port-type    Power-state
  ---------------------------------------------------------------------------
       1    ETH          No
       2    ETH          Yes
       3    ETH          No
       4    ETH          Yes
  ---------------------------------------------------------------------------
  Whether the reversed PoE has local power supply: Yes
  Note: Local power supply of the device supporting reversed PoE is used for
        only commissioning and it cannot be used for normal power supply
  ---------------------------------------------------------------------------
```

- Reverse PoE work mode indicates the reversed PoE work mode.
- Power-state indicates reverse power supply status. **Yes** indicates that reversed PoE exists on a port. **No** indicates that reversed PoE does not exist on a port. **-** indicates that reversed PoE is not supported on a port.
- For Whether the reversed PoE has local power supply, **Yes** indicates that the local power supply exists, **No** indicates that the local power supply does not exist, and **-** indicates that reversed PoE is not supported. The local power supply of the reversed PoE is used for commissioning only. It cannot be used as the working power supply.

Step 7 (Optional) Configure ONU maintenance measures.

1. Configure ONU automatic migration.

   When the ONU automatic migration function is enabled, if a PON port or fiber is faulty, the ONU connected to this faulty PON port is migrated to another PON port. If the ONU automatic adding policy is met, the ONU is automatically removed from the original PON port, and then added to the new PON port. In this process, the OLT does not report the SN authentication conflict alarm. The ONU connected to the original PON port must meet the following conditions:

   – The ONU is offline.
   – The ONU supports OMCI management.
   – The ONU is not a member of the type C protection group.
   – The ONU supports SN authentication.

   ```
   huawei(config)#ont auto-migration enable
   ```

2. Configure ONU automatic aging.

   When an offline ONU is aged and meets automatic adding policy, the OLT automatically deletes the aged ONU to save resources. The ONU automatic aging period can be configured. The value range is 1–365. The default value is

 Copyright © Huawei Technologies Co., Ltd.

no-aging, which indicates no aging. For an aged ONU that meets the automatic adding policy, it will be automatically registered after being powered on.

```
huawei(config)#ont auto-aging 15
```

3. Run the **display ont global-config** or **display xpon global-config** to check whether the ONU automatic migration or aging function is enabled.

```
huawei(config)#display ont global-config
  ------------------------------------------------------------------------
  ......
  ONT auto-migration switch                 : enable
  ONT auto-aging time(day)                  : 15
  ------------------------------------------------------------------------
```

Step 8 Add an ONU.

1. Enable the ONU automatic discovery function on a GPON port.

   When the ONU automatic adding policy is configured, the ONU does not need to be added in the offline mode, because it will be automatically found after it goes online.

```
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#port 0 ont-auto-find enable
```

2. (Optional) Confirm an ONU.

   During ONU automatic adding, if **auto-confirm** is **disable**, run the **ont-confirm** command to confirm the ONU. When the **ont confirm** command is run and no line profile or service profile is specified, if the ONU type is matched with that ONU type configured in the automatic adding policy, the ONU is bound to the line and service profiles configured in the automatic adding policy. If the ONU type is not matched, the ONU is bound to default line profile 0.

```
huawei(config-if-gpon-0/1)#display ont autofind 0     /This command is run to display information about all ONTs connected to the GPON port through the optical splitter.
huawei(config-if-gpon-0/1)#ont confirm 0 ontid 1 sn-auth 3230313126595540 omci
```

3. (Optional) Change the ONU description.

   Run the **ont modify** *portid ontid* **desc** *describe-value* command to change the ONU description. It is recommended that information about the location and time be added for each ONU to facilitate the fault locating and maintenance.

```
huawei(config-if-gpon-0/1)#ont modify 0 1 desc ftto_2015
```

4. (Optional) Bind an alarm threshold profile.

   In this example, the default profile (profile 1) is used.

```
huawei(config-if-gpon-0/1)#ont alarm-profile 0 1 profile-id 1
```

5. Check ONU status.

   After the ONU is added, run the **display ont info** command to query current ONU status. Ensure that Control flag is **active**, Run State is **online**, Config state is **normal**, and Match state is **match**.

```
huawei(config-if-gpon-0/1)#display ont info 0 1
  ------------------------------------------------------------------------
  F/S/P               : 0/1/0
  ONT-ID              : 1
  Control flag        : active    //Indicates that the ONU is activated.
  Run state           : online    //Indicates that the ONU is online.
```

 Copyright © Huawei Technologies Co., Ltd.

```
  Config state         : normal    //Indicates that the ONU configuration state is normal.
  Match state          : match     //Indicates that the capability profile bound to the ONU is consistent
with actual ONU capability.
...//
```

When **Config state** is **failed** and **Run state** is **offline**,

- If **Control flag** is **deactive**, run the **ont activate** command to activate the ONU in the GPON port mode.
- If **Run state** is **offline**, a physical line may be disconnected or the optical module may be damaged. Check the line and the optical module.
- If **Config state** is **failed**, the configured ONU capability exceeds the actual ONU capability. In this case, run the **display ont failed-configuration** command in the diagnose mode to check the failed configuration item and the failure cause. Then, rectify the fault accordingly.

Run the **display ont info summary** command to query ONU information about the SN, type, operating status, online time, offline time, and offline cause.

Step 9 Query the configuration status of a service flow.

After the MA5671-FE24 is powered on and connected to the PON, the service flow is automatically created after about 1 minute. Run the **display service-port all** command to check whether the service flow is successfully configured.

```
huawei(config)#display service-port all
 Switch-Oriented Flow List
 -----------------------------------------------------------------------------
  INDEX VLAN VLAN      PORT F/ S/ P VPI  VCI   FLOW  FLOW       RX   TX   STATE
        ID   ATTR      TYPE                TYPE  PARA
 -----------------------------------------------------------------------------
      0  100 common   gpon 0/1 /0  1    1     vlan  1          7    7    up
      1  100 common   gpon 0/1 /0  1    1     vlan  2          7    7    up
     ...
      8  100 common   gpon 0/1 /0  1    2     vlan  9          7    7    up
      9  100 common   gpon 0/1 /0  1    2     vlan  10         7    7    up
     ...
     16  100 common   gpon 0/1 /0  1    3     vlan  17         7    7    up
     17  100 common   gpon 0/1 /0  1    3     vlan  18         7    7    up
    ...
```

The main parameters are described as follows:

- **VPI** indicates the ONU ID.
- **VCI** indicates the GEM port ID.
- **FLOW PARA** indicates the user VLAN.

When a service flow with double-tagged VLAN is created, run the **display service-port** *index* command to query the inner VLAN on the network side. For example, for a service flow which is automatically created, the start VLAN is 2000, the ONU ID is 1, the number of ONU ports is 24, and the ONU user port ID is 1, the query results are shown in the following. Among them, **Label** indicates the inner VLAN on the network side.

```
huawei(config)#display service-port 0
 ------------------------------------------------------------
 Index              : 0
 VLAN ID            : 100
 VLAN attr          : QinQ
 Port type          : gpon
```

 Copyright © Huawei Technologies Co., Ltd.

```
F/S/P              : 0/1/0
ONT ID             : 1
GEM port index     : 1
ONT Port           : 1
Flow type          : vlan
Flow para          : 1
TX                 : 7
Inbound table name  : ftto_hsi
RX                 : 7
Outbound table name : ftto_hsi
Admin status       : enable
State              : up
Label              : 2025
Priority           : 0
PVC bundle         : no
Max MAC count      : 1023
Tag transform      : translate-and-add
Description        :
Remote description  :
Service-port bundle : -
Cos                : -
Car-Group          : -
Static MAC         :
IP address         :
-------------------------------------------------------------
```

----End

10.3.4 Maintenance and Diagnosis

A service flow is a Layer 2 forwarding channel between an access node and a user. Check the status of a service flow using the following methods:

- After a user goes online, run the **display mac-address** command to query the MAC address of the user. This method is used to check whether the configuration of the service flow is correct.
- After a service is provisioned, run the **display statistics service-port** command to query the packet statistics of the service flow or run the **display traffic service-port** command to query the packet transmit and receive rate of the service flow in real time. The query results can be used to detect a service exception or locate a fault.

10.4 Service Port Bundle

10.4.1 What Is Service Port Bundle

Definition

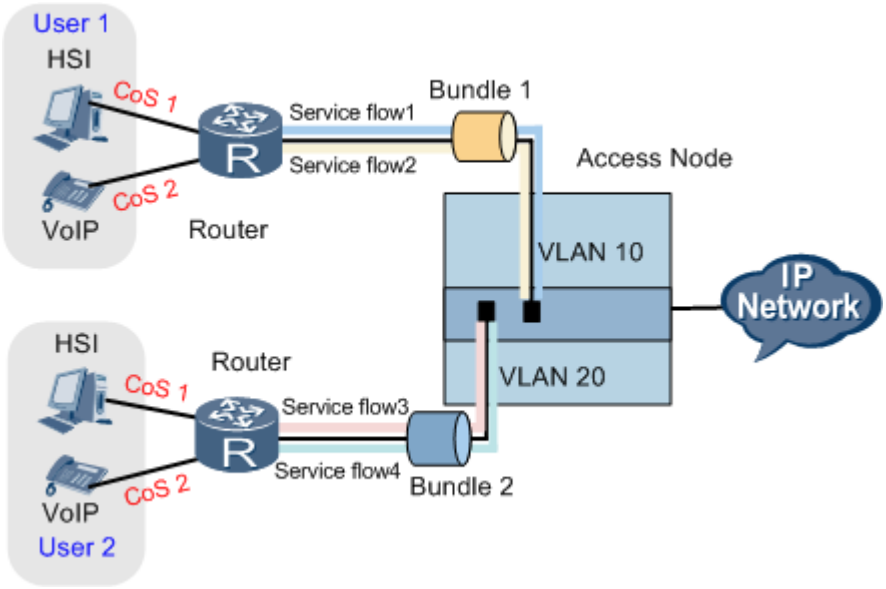
In a network where VLANs are planned based on ports on the access node and the access node connects to users through a router, the packets transmitted from a user to the access node through the router may carry the same MAC address even when they carry different services. As a result, MAC address transfer may occur between different service flows (service flow 1 and service flow 2) on the access node. The service port bundle feature addresses this issue by using CoS-based route selection.

Service port bundle, also named service flow bundle or flow bundle, is a CoS-based packet forwarding model. Each service port bundle corresponds to a group

 Copyright © Huawei Technologies Co., Ltd.

of services for a user. Each service flow carries one type of service and has a CoS level. A service port bundle can also be considered a bundle of service flows. The following figure shows the schematic diagram for service port bundle.

Figure 10-25 Schematic diagram for service port bundle



Benefits

Benefits to carriers

CoS identification is introduced to the service port bundle principle. Consequently, service port bundle breaks the VLAN+MAC and S-VLAN+C-VLAN forwarding policies for the access node down to the VLAN+MAC+CoS and S-VLAN+C-VLAN +CoS forwarding policies and improves the flexibility in service planning for carriers.

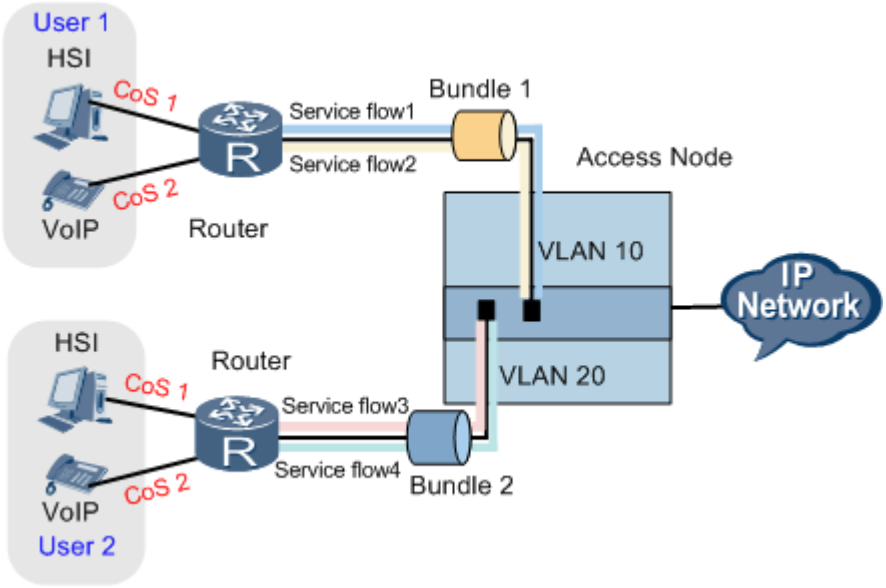
10.4.2 Schematic Diagram For Service Port Bundle

The following explains the schematic diagram for service port bundle.

- The access node connects to users through a router. CoS levels are used to distinguish different services of a user.
- VLANs on the access node are planned based on ports. Services of different users are carried by different service flows. Service flows of a user are bundled together, called the service port bundle, or service flow bundle.

 Copyright © Huawei Technologies Co., Ltd.

Figure 10-26 Schematic diagram for service port bundle



Upstream direction

The Access node receives packets of the service port bundle, and forwards the received packets to the upper-layer network through upstream ports in the S-VLAN (in the above figure, service flows 1 and 2 in service port bundle 1 are converged to the upstream port in VLAN 10). At the same time, the Access node obtains the service port bundle identifier from the ARL table (learned from the user port) and creates an entry for the service port bundle.

Downstream direction

The Access node determines the user port based on S-VLAN+MAC or S-VLAN+C-VLAN, queries the service port bundle table, determines a service flow based on the CoS value, and forwards the received packets through the service flow. In this way, the Access node implements the VLAN+MAC+CoS or S-VLAN+C-VLAN+CoS forwarding policy.

For more information about Layer 2 forwarding policies, see **10.5 Layer 2 Forwarding Policy**.

10.4.3 Configuring a Service Port Bundle

On an access node, VLANs are planned based on ports. If the access node connects to users through a router, configure a service port bundle to prevent MAC address flapping between various service flows of the users.

Procedure

Step 1 Run the **service-port-bundle** command to create a service port bundle.

A service port bundle corresponds to a group of services. It is recommended that you plan a unique service port bundle ID for a user.

Run the **service-port-bundle** command to create a service port bundle.

 Copyright © Huawei Technologies Co., Ltd.

Step 2 Add service flows to the service port bundle and set classes of service (CoSs) for the service flows.

The services in a service flow bundle are carried by service flows. Add the service flows to the service port bundle and set a CoS for each service flow. In this way, the router forwards service packets according to the CoS.

Run the **service-port** command to set **bundle** *bundleid* **cos** *cos* when creating a service flow. *bundleid* is the ID of the service port bundle created in the preceding step.

NOTE

- A service port bundle implements packet forwarding based on CoS levels, meeting the QoS demands of the upper-layer network. Hence, it requires that packets from the upper-layer network devices to the access device contain correct CoS values. The CoS value of a service that is defined by the service port bundle at the access node must be consistent with the CoS value defined by the upper-layer network.
- For users in a service port bundle, one user goes online, updates the lease, and goes offline through only one service flow, and the CoS value of upstream packets must be consistent with that of downstream packets.

----End

Example

The following is an example of the configurations used to configure a service port bundle to prevent MAC address flapping between Internet access and multicast service flows of user A:

- GPON port 0/2/1 on the access node connects to user A .
- The VLANs of the access node are planned based on ports.
- GPON port 0/2/1 is in VLAN 10.
- User A requires the Internet access and broadcast services.
- The ID of the service port bundle for user A is 10.
- The Internet access and broadcast services of user A are differentiated based on CoS values.
- For the Internet access service:
  - The CoS value is 1.
  - The ID of the traffic profile is 10.
- For the broadcast service:
  - The CoS value is 3.
  - The ID of the traffic profile is 20.

```
huawei(config)#service-port-bundle 10
huawei(config)#service-port 10 vlan 10 eth 0/2/1 multi-service user-vlan 10 bundle 10 cos 1 inbound traffic-table index 10 outbound traffic-table index 10
huawei(config)#service-port 11 vlan 10 eth 0/2/1 multi-service user-vlan 20 bundle 10 cos 3 inbound traffic-table index 20 outbound traffic-table index 20
```

10.5 Layer 2 Forwarding Policy

 Copyright © Huawei Technologies Co., Ltd.

10.5.1 Layer 2 Forwarding Policy Overview

The MA5800, as Layer 2 network equipment, supports the ability to transparently transmit or forward packets at Layer 2. Traditionally, Layer 2 packet forwarding is based on VLANs and MAC addresses of packets (VLAN+MAC). If the destination MAC address of a packet becomes ineffective due to dynamic MAC address aging, VLAN+MAC searching fails. The packet becomes an unknown unicast packet and is broadcast within the VLAN, which poses a security threat. In addition, VLAN+MAC forwarding is subject to MAC spoofing and attacks, which lead to security problems.

To address the preceding problems, you can use S-VLAN+C-VLAN forwarding instead.

In S-VLAN+C-VLAN forwarding, two VLAN tags form a Layer 2 forwarding mapping relationship. Packets are forwarded based on VLANs rather than MAC addresses.

10.5.2 Layer 2 Forwarding Policy Principles

The MA5800 supports VLAN+MAC forwarding, S-VLAN+C-VLAN forwarding, and two other forwarding modes derived from the first two: VLAN+MAC+CoS and S-VLAN+C-VLAN+CoS.

VLAN+MAC Forwarding

With the VLAN+MAC forwarding policy, the system automatically learns the mapping between the VLAN, source MAC address, and incoming port when packets enter the system. According to the VLAN and destination MAC address, the system then searches for the outgoing port and transmits the packets through this port.

In the VLAN+MAC forwarding policy, if the source MAC address is a broadcast or unicast MAC address, packets are broadcast in the VLAN. That is, packets are duplicated and transmitted to every port in the VLAN.

S-VLAN+C-VLAN Forwarding

The two VLANs (S-VLAN and C-VLAN) are an extension of VLAN. This expands the VLAN ID range. In addition, S and C have different meanings. S stands for service, and C stands for customer (user). Each S-VLAN+C-VLAN uniquely identifies a user service, and S-VLAN+C-VLAN forwarding can be implemented.

In S-VLAN+C-VLAN forwarding, two VLAN tags form a Layer 2 forwarding mapping relationship to implement VLAN-based forwarding. S-VLAN+C-VLAN forwarding is classified into non-strict S-VLAN+C-VLAN forwarding and strict S-VLAN+C-VLAN forwarding. Strict S-VLAN+C-VLAN forwarding entries do not need to be learned dynamically. The system automatically creates static forwarding entries during the establishment of service flows. Non-strict S-VLAN+C-VLAN forwarding entries need to be dynamically learned. According to the forwarding entries, upstream packets are transmitted through the corresponding upstream port and downstream packets are transmitted through the corresponding service port.

 Copyright © Huawei Technologies Co., Ltd.

Figure 10-27 Strict S-VLAN+C-VLAN forwarding model

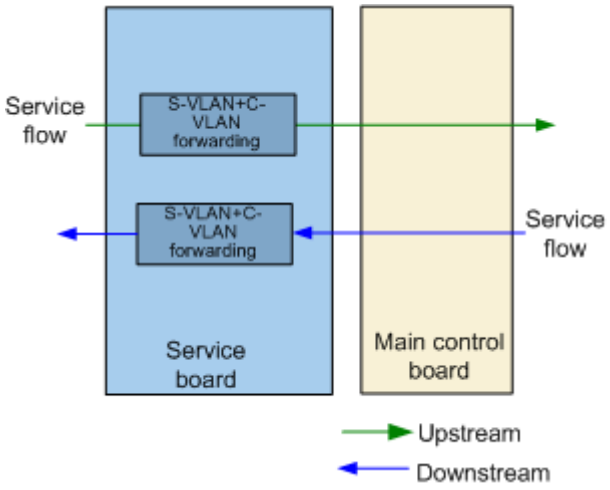
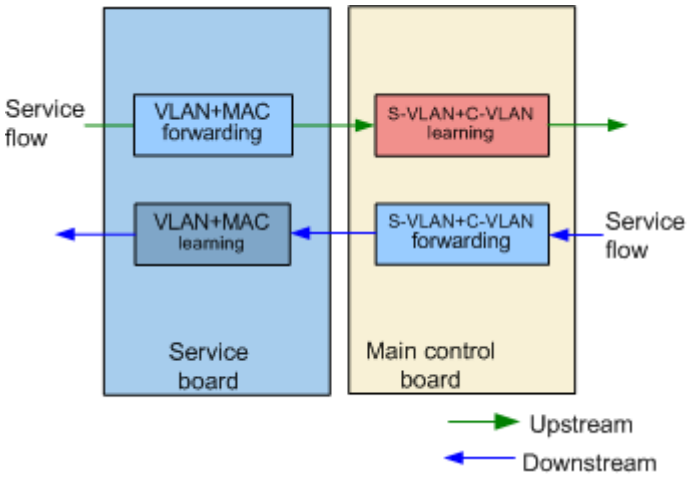


Figure 10-28 Non-strict S-VLAN+C-VLAN forwarding model



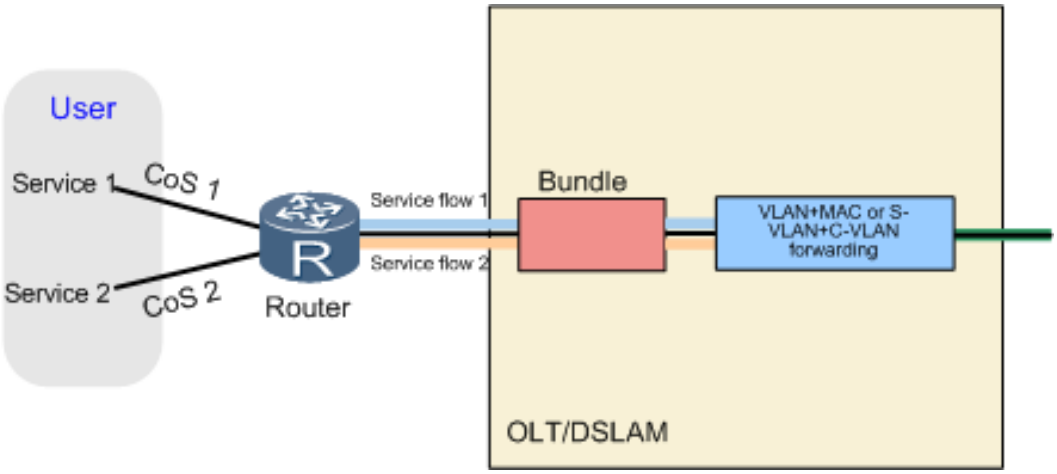
VLAN+MAC+CoS and S-VLAN+C-VLAN+CoS Forwarding

VLAN+MAC+CoS and S-VLAN+C-VLAN+CoS forwarding policies are derived from VLAN+MAC and S-VLAN+C-VLAN forwarding policies of the MA5800 respectively after class of service (CoS) identification is introduced based on service port bundles. The forwarding policies based on CoS resolve the problem of different service flows with the same VLAN+MAC or S-VLAN+C-VLAN and therefore improve service planning flexibility.

Figure 10-29 shows the CoS-based forwarding model.

 Copyright © Huawei Technologies Co., Ltd.

Figure 10-29 VLAN+MAC+CoS and S-VLAN+C-VLAN+CoS forwarding model



Upstream direction

All service flows from the same user converge to a service port bundle. The MA5800 learns the service port bundle identifier of the service flows, creates a service port bundle entry, and performs Layer 2 packet forwarding based on VLAN +MAC or S-VLAN+C-VLAN.

Downstream direction

The OLT determines a unique user based on VLAN+MAC or S-VLAN+C-VLAN, queries the service port bundle entries, finds the service flows based on CoS, and forwards packets.

10.5.3 Configuring a Layer 2 Forwarding Policy

Layer 2 forwarding policy configuration is the basis of VLAN configuration and an important step to ensure correct service forwarding. Before configuring services, ensure that the VLAN forwarding policy has been configured as planned.

Prerequisites

- VLAN IDs have been created correctly.
- VLAN attributes have been set correctly.

Context

A VLAN forwarding policy is a type of Layer 2 forwarding policy. Two types of VLAN forwarding policy are available: **vlan-connect** or **vlan-mac**. **vlan-mac** indicates VLAN+MAC forwarding and **vlan-connect** indicates S-VLAN+C-VLAN forwarding.

To implement VLAN+MAC+CoS or S-VLAN+C-VLAN+CoS forwarding, configure the VLAN forwarding policy and service port bundles separately.

Default Configuration

Table 10-27 lists the default VLAN forwarding policy settings.

 Copyright © Huawei Technologies Co., Ltd.

Table 10-27 Default VLAN forwarding policy settings

Parameter	Default Value
forwarding-mode	VLAN+MAC

Procedure

- Configure VLAN+MAC forwarding.

  You can configure a VLAN forwarding policy using two methods. Select an appropriate one as follows:

  – To configure a VLAN forwarding policy for a single VLAN, use the first method.

  – To configure the same forwarding policy for multiple VLANs with the same service profile parameters, use the second method (batch configuration using a service profile).

  – Method 1

  a. In global configuration mode, run the **vlan forwarding** command.

  – Method 2

  a. Run the **vlan service-profile** command to create a VLAN service profile and enter the VLAN service profile mode.

  b. Run the **forwarding** command to configure a VLAN forwarding policy.

  c. Run the **commit** command for the profile configuration to take effect.

  d. Run the **quit** command to quit the VLAN service profile mode.

  e. Run the **vlan bind service-profile** command to bind a VLAN to the VLAN service profile.

- Configure S-VLAN+C-VLAN forwarding.

  a. Run the **vlan** command to add a VLAN.

  b. Run the **vlan attrib** command to change the VLAN attribute to QinQ or stacking.

  c. Run the **vlan forwarding** or **forwarding** command to configure a VLAN forwarding policy.

  d. Run the **port vlan** command to associate upstream ports with S-VLANs and C-VLANs.

  e. Run the **service-port** command to create a service port.

----End

Result

After the configuration, service flows can be created based on the configured VLANs and users connected to the MA5800 can ping the upstream equipment.

Example

Example: Configure the forwarding policy of VLAN 50 as S-VLAN+C-VLAN.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#vlan forwarding 50 vlan-connect
```

Example: Configure the forwarding policy of VLAN 65 as strict S-VLAN+C-VLAN. VLAN 73 indicates an enterprise VLAN on which MAC address learning is disabled.

```
huawei(config)#vlan 65 smart
huawei(config)#vlan attrib 65 stacking
huawei(config)#vlan service-profile profile-id 200
huawei(config-vlan-srvprof-200)#mac-address learning fabric disable
huawei(config-vlan-srvprof-200)#forwarding vlan-connect
huawei(config-vlan-srvprof-200)#commit
huawei(config-vlan-srvprof-200)#quit
huawei(config)#vlan bind service-profile 65 profile-id 200
huawei(config)#port vlan 65 inner-vlan-list 73 0/9 0
huawei(config)#service-port 100 uplink-port 0/9/0 vlan 65 eth 0/2/1 multi-service user-vlan 73 rx-cttr 10 tx-cttr 10
```

10.6 Layer 2 User Bridging

Layer 2 user bridging implements Layer 2 data exchange among users under the same MA5800.

10.6.1 Layer 2 User Bridging Overview

Users under the same MA5800 are isolated at Layer 2, but QinQ service deployment requires users communicate with each other at Layer 2. Therefore, QinQ services can be deployed only between different MA5800s.

If users in the same IP network segment on the same MA5800 need to exchange data, upstream equipment should support ARP proxy to implement Layer 3 data forwarding.

The Layer 2 user bridging feature addresses the preceding issues. After Layer 2 user bridging is enabled, users under the same MA5800 can exchange data at Layer 2.

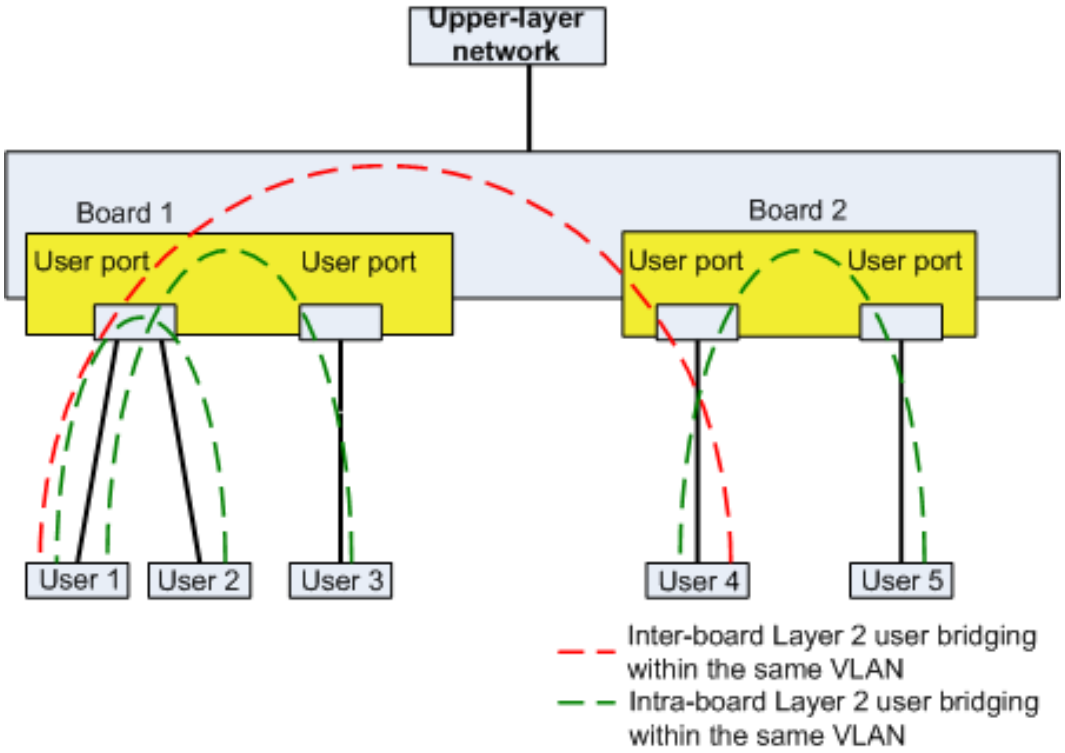
10.6.2 Layer 2 User Bridging Principles

OLT supports VLAN-based Layer 2 user bridging.

 Copyright © Huawei Technologies Co., Ltd.

VLAN-based Layer 2 User Bridging

Figure 10-30 VLAN-based Layer 2 user bridging model

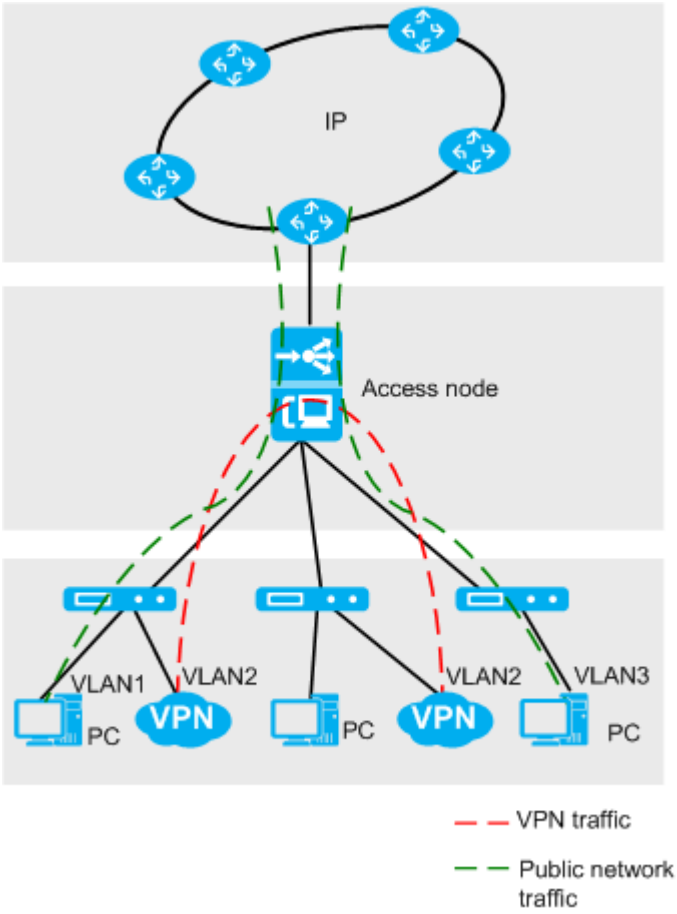


As shown in **Figure 10-30**, the following functions are implemented on the MA5800 after Layer 2 user bridging is enabled on a VLAN.

- Inter-board user bridging: bridging among ports on different boards, for example, users 1, 2, and 3 can exchange data with users 4 and 5
- Intra-board user bridging:
  - Bridging among different users under the same port, for example, users 1 and 2 can exchange data with each other
  - Bridging among users under different ports on the same board, for example, users 1, 2, and 3 can exchange data
  - Bridging among different ports on the same board, for example, users 4 and 5 can exchange data with each other

 Copyright © Huawei Technologies Co., Ltd.

Figure 10-31 VLAN-based user bridging network



As shown in **Figure 10-31**, on one MA5800, Layer 2 bridging is implemented among enterprise private line users (VPN represented by the red line) but is not implemented among some common access users (represented by the green line). Services are classified by VLAN. VLAN 2 is a QinQ VLAN for carrying enterprise private line services. Enable Layer 2 user bridging on VLAN 2 so that private line users on the same MA5800 can exchange data.

10.6.3 Configuring VLAN-based Layer 2 User Bridging

Default Setting

By default, Layer 2 user bridging based on VLAN service profiles is disabled.

Procedure

Step 1 In global configuration mode, run the **vlan service-profile** command to enter the VLAN service profile mode.

Step 2 Run the **user-bridging** command to enable Layer 2 user bridging on a VLAN service profile.

Step 3 Run the **commit** command for the settings to take effect.

 Copyright © Huawei Technologies Co., Ltd.

Step 4 Go back to the global configuration mode and run the **vlan bind service-profile** command to bind a VLAN to the VLAN service profile.

NOTE

- If the VLAN has been bound to the VLAN service profile, you do not need to bind the VLAN again.
- When the users of the same OLT are configured with Layer 2 intercommunication in the VLAN domain and do not need to communicate with users of other networks outside the OLT, you do not need to configure the VLAN tag on the upstream port.

----End

Result

If you run the **display vlan service-profile** command, **User-bridging** is **enable** for the VLAN service profile. Then, under the MA5800, users in the same VLAN with Layer 2 user bridging enabled can be pinged with each other

Example

Example: VLAN 10 is the S-VLAN for enterprise private line services. The VLAN is bound to VLAN service profile 3. Run the following commands to enable Layer 2 user bridging on VLAN 10 so that users in the VLAN on the same MA5800 can communicate with each other at Layer 2.

```
huawei(config)#vlan service-profile profile-id 3
huawei(config-vlan-srvprof-3)#user-bridging enable
huawei(config-vlan-srvprof-3)#commit
huawei(config-vlan-srvprof-3)#quit
```

ONU1 and ONU2 are deployed in enterprise branches and connect to the OLT through aggregation links. ONU1 and ONU2 can communicate with each other at Layer 2. The upstream port of the OLT is 0/9/0, the VLAN ID of the L2 intercommunication service is 300, and the GEM Port ID is 12. Create a VLAN service profile, enable the VLAN Layer 2 intercommunication function, and configure the protocol transparent transmission attributes related to the Layer 2 intercommunication service.

```
huawei(config)#vlan 300 smart
huawei(config)#port vlan 300 0/9 0
huawei(config)#traffic table ip index 10 cir 102400 priority user-cos 4 priority-policy tag-in-package
huawei(config)#service-port 301 vlan 300 gpon 0/2/0 ont 1 gemport 12 multi-service user-vlan 300 rx-cttr 10 tx-cttr 10
huawei(config)#service-port 302 vlan 300 gpon 0/2/1 ont 2 gemport 12 multi-service user-vlan 300 rx-cttr 10 tx-cttr 10
huawei(config)#vlan service-profile profile-id 1
huawei(config-vlan-srvprof-1)#user-bridging enable  // Mandatory
huawei(config-vlan-srvprof-1)#bpdu tunnel enable
huawei(config-vlan-srvprof-1)#ospf tunnel enable
huawei(config)#vlan bind service-profile 300 profile-id 1
```

 Copyright © Huawei Technologies Co., Ltd.

11 Layer 3 Forwarding

This topic describes the network layer (Layer 3) features implemented by the system.

11.1 IPv4

11.2 IPv6
Internet Protocol Version 6 (IPv6), also called IP Next Generation (IPng), is a second-generation protocol of the network layer protocol family.

11.3 ARP

11.4 ARP Proxy
ARP proxy is used to communication between users in different VLAN or in the same VLAN but isolated from each other by layer 3 forwarding.

11.5 DHCP Server

11.6 DHCPv6 Server

11.7 DHCP Relay
This section describes the implementation and configuration of Dynamic Host Configuration Protocol (DHCP) relay.

11.8 DHCP Proxy
This section describes the implementation and configuration of Dynamic Host Configuration Protocol (DHCP) proxy.

11.9 BFD

11.10 TWAMP
The Two-Way Active Measurement Protocol (TWAMP) is a technology that measures the round-trip performance of an IP network.

11.11 TWAMP Light
Two-Way Active Measurement Protocol (TWAMP) Light is a light version of TWAMP, which is defined in RFC 5357. TWAMP Light measures the round-trip performance of an IP network by using simplified control protocols to establish test sessions.

11.12 Routing
Routing is a common term used for describing the path through which the packets from a host in a network travel to a host in another network.

 Copyright © Huawei Technologies Co., Ltd.

11.1 IPv4

11.1.1 Introduction to IPv4

Definition

At the core of the TCP/IP protocol suite, Internet Protocol Version 4 (IPv4) works at the Internet layer in the TCP/IP model. This layer corresponds to the network layer in the OSI model. At the IP layer, information is divided into data units, and address and control information is added to allow datagrams to be routed.

IP provides unreliable and connectionless data transmission services. Unreliable transmission means that IP does not ensure that IP datagrams successfully arrive at the destination. IP only provides best effort delivery. Once an error occurs, for example, a router exhausts the buffer, IP discards the excess datagrams and sends ICMP messages to the source. The upper layer protocols, such as TCP, are responsible for resolving reliability issues.

Connectionless transmission means that IP does not maintain status information for subsequent datagrams. Every datagram is processed independently, meaning that IP datagrams may not be received in the same order they are sent. If a source sends two consecutive datagrams A and B in sequence to the same destination, each datagram is possibly routed over a different path to the destination, and therefore B may arrive ahead of A.

Purpose

IPv4 shields link layer protocol differences and provides a uniform standard for transmission at the network layer.

11.1.2 Principles to IPv4

11.1.2.1 ICMP

The Internet Control Message Protocol (ICMP) is an error-reporting mechanism and is used by IP or an upper-layer protocol (TCP or UDP). An ICMP message is encapsulated as a part of an IP datagram and transmitted through the Internet.

An IP datagram contains information about only the source and destination, not about all nodes along the entire path through which the IP datagram passes. The IP datagram can record information about all nodes along the path only when route record options are set in the IP datagram. Therefore, if a device detects an error, it reports the error to the source and not to intermediate devices.

When an error occurs during the IP datagram forwarding, ICMP reports the error to the source of the IP datagram, but does not rectify the error or notify the intermediate devices of the error. A majority of errors generally occur on the source. When an error occurs on an intermediate device, however, the source cannot locate the device on which the error occurs even after receiving the error report.

 Copyright © Huawei Technologies Co., Ltd.

Redirect Message

After a device receives an IP datagram that should have been sent to another device, the device sends an ICMP Redirect message to the sending device to redirect the IP datagram. Redirect messages help hosts that have a limited amount of routing information to gradually build up complete routing tables. A host may initially have only one default routing entry in the routing table. When an error occurs in the default routing entry, the default gateway notifies the host of generating another route to redirect the IP datagram and allows the host to modify the routing table.

ICMP Redirect messages can be generated and sent when all of the following conditions are met:

- The IP datagram is received and forwarded by the same interface. Note that both the physical interface and VLANIF interface are the same.
- The route used to send an IP datagram is not generated or modified through an ICMP Redirect message. That is, when a redirected route is used to send an IP datagram, no ICMP Redirect message is generated when the IP datagram is received.
- The route through which an IP datagram is transmitted cannot be the default route.
- The subnet mask of the source address in an IP datagram is the same as the subnet mask of the outgoing interface's address.

A large number of generated redirection ICMP packets will degrade the device performance. The device does not send redirection ICMP packets by default. The **ip redirect enable** command is used to enable the function of sending redirection ICMP packets.

TTL timeout message

During the process of forwarding or assembling an IP datagram, if the time-to-live (TTL) field in the IP datagram is zero, the receiving device sends a TTL timeout message to the source.

Port Unreachable Message

If a host or routing device receives a local UDP or TCP datagram but cannot find the corresponding process according to the destination port of the datagram, the host or routing device sends a Port Unreachable message to the source.

Destination Unreachable Message

If a network is unreachable, route selection fails. If a host is unreachable, message delivery fails. The source device can determine which address is unreachable by checking the IP header and the leftmost 64 bits in the original IP datagram (Internet Header + 64 bits of the Original Data Datagram field).

When a routing device forwards a message that meets the following conditions:

- No route is available for the destination address of the message.
- The message is not sent for the local host.

 Copyright © Huawei Technologies Co., Ltd.

The routing device will discard the message and return an ICMP Host Unreachable message to the source address to inform the source host to stop sending messages to this destination.

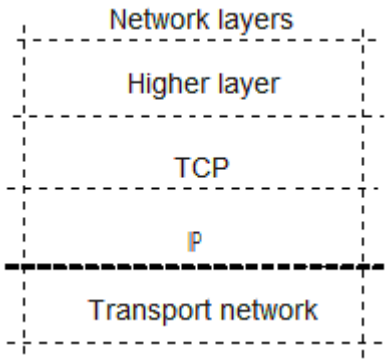
11.1.2.2 TCP

The Transmission Control Protocol (TCP) defined in RFC793 ensures high-reliability transmission between hosts. TCP provides reliable, connection-oriented, and full-duplex services for user processes. TCP transmits data through sequenced and nonstructural byte streams.

TCP is an end-to-end, connection-oriented, and reliable protocol. TCP supports multiple network applications. In addition, TCP assumes that the lower layer provides only unreliable datagram services, and it can run over a network of different hardware structures.

Figure 11-1 shows the position of TCP in a layered protocol architecture, where TCP is above IP. TCP can transmit variable-length data through IP encapsulation. IP then performs data fragmentation and assembly and transmits the data over multiple networks.

Figure 11-1 TCP in the layered protocol architecture



TCP works below applications and above IP. Its upper-layer interface consists of a series of calls similar to the interrupt call of an operating system.

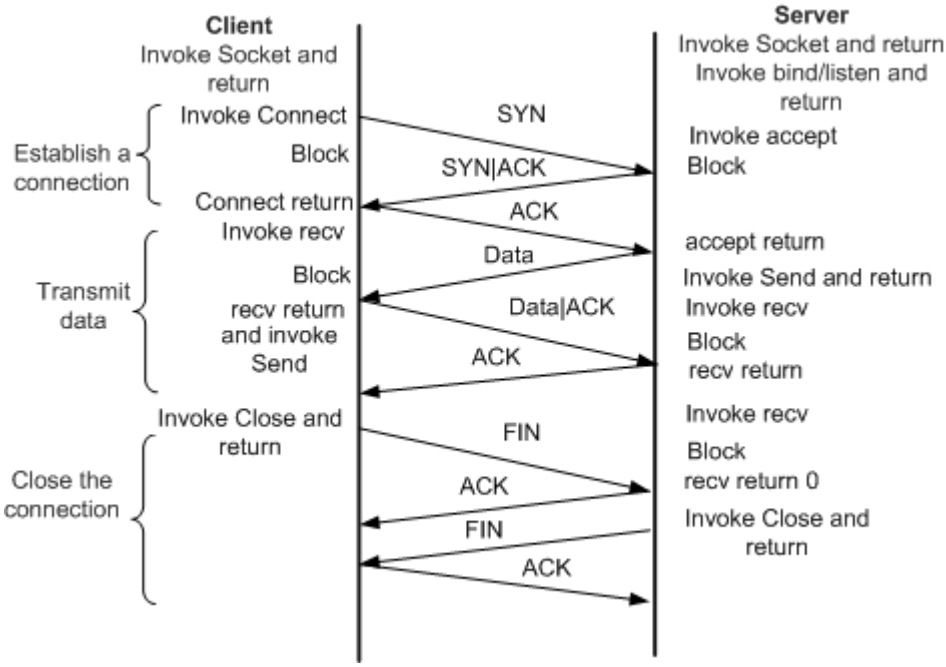
TCP can asynchronously transmit data of upper-layer applications. The lower-layer interfaces are assumed as IP interfaces. To implement connection-oriented and reliable data transmission over unreliable networks, TCP must provide the following functions:

- Ensures reliability and flow control.
- Provides multiple interfaces for upper-layer applications.
- Provides data for multiple applications.
- Ensures connections.
- Ensures communication security.

Figure 11-2 shows the process of setting up and tearing down a TCP connection.

 Copyright © Huawei Technologies Co., Ltd.

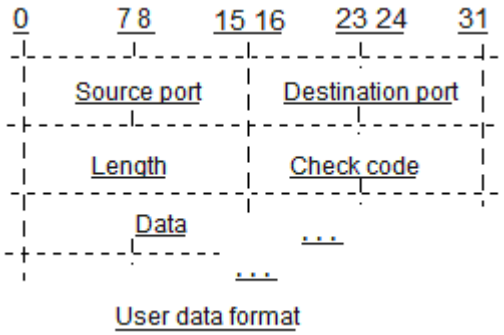
Figure 11-2 Setup and teardown of a TCP connection



11.1.2.3 UDP

The User Datagram Protocol (UDP) is a computer communication protocol that provides packet switching services on the Internet. By default, UDP uses IP as the lower-layer protocol. UDP provides the simplest protocol mechanism that sends information to a user application. UDP is transaction-oriented and does not support delivery or duplicate protection. TCP, however, is required by applications for reliable data transmission. **Figure 11-3** shows the format of a UDP datagram.

Figure 11-3 UDP datagram format



11.1.2.4 RawIP

RawIP only fills in certain fields of an IP header and allows an application to provide its own IP header. Similar to UDP, RawIP is unreliable. No control

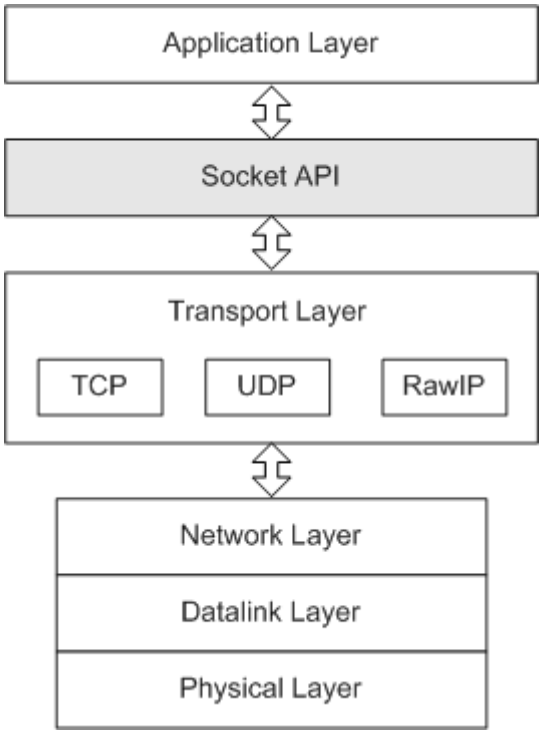
 Copyright © Huawei Technologies Co., Ltd.

mechanism is available to verify whether a RawIP datagram is received. RawIP is connectionless, and it transmits data between hosts without an electric circuit of any type. Unlike UDP, RawIP allows application data to be directly processed at the IP layer through a socket. This is helpful to the applications that need to directly communicate with the IP layer.

11.1.2.5 Socket

A socket consists of a set of application programming interfaces (APIs) working between the transport layer and application layer. The socket shields differences of transport layer protocols and provides the uniform programming interfaces for the application layer. In this manner, the application layer, being exempt from the detailed process of the TCP/IP protocol suite, can transmit data over IP networks by calling socket functions. **Figure 11-4** shows the position of the socket in the TCP/IP protocol stack.

Figure 11-4 Schematic diagram of the socket in the TCP/IP protocol stack



The following types of sockets are supported by different protocols at the transport layer:

- TCP-based socket: provides reliable byte-stream communication services for the application layer.
- UDP-based socket: supports connectionless and unreliable data transmission for the application layer and preserves datagram boundaries.
- RawIP socket: also called raw socket. Similar to the UDP-based socket, the RawIP socket supports connectionless and unreliable data transmission and preserves datagram boundaries. The RawIP socket is unique in that it can be used by applications to directly access the network layer.

 Copyright © Huawei Technologies Co., Ltd.

- Link layer-based socket: used by Intermediate System to Intermediate System (IS-IS) to directly access the link layer.

11.1.3 IPv4 Reference Standards and Protocols

The following table lists the references.

Document No.	Document Name
RFC 793	Transmission Control Protocol
RFC 768	User Datagram Protocol
RFC 1155	Structure and identification of management information for TCP/IP-based internets
RFC 1200	IAB official protocol standards
RFC 1219	On the assignment of subnet numbers
RFC 1878	Variable Length Subnet Table For IPv4
RFC 2581	TCP Congestion Control
RFC 2644	Changing the Default for Directed Broadcasts in Routers
RFC 5925	The TCP Authentication Option

11.2 IPv6

Internet Protocol Version 6 (IPv6), also called IP Next Generation (IPng), is a second-generation protocol of the network layer protocol family.

11.2.1 Why IPv6 is Required

Definition

Internet Protocol Version 6 (IPv6), complies with a set of specifications defined by the Internet Engineering Task Force (IETF), IPv6 is an upgrade of Internet Protocol Version 4 (IPv4). The most significant difference between IPv6 and IPv4 is that the length of IP addresses is extended from 32 bits to 128 bits in IPv6. Featuring a simplified header format, sufficient address space, layered address structure, flexible extension header, and enhanced neighbor discovery (ND) mechanism, IPv6 is more competitive in the future market.

Purpose

As the IPv4-based Internet achieves great success, the IP technology is widely applied. With the rapid development of the Internet, however, deficiencies of IPv4 are more strongly felt, especially in the following aspects:

 Copyright © Huawei Technologies Co., Ltd.

- The IPv4 address space is insufficient.

  An IPv4 address is identified by using 32 bits. In theory, a maximum of 4.3 billion addresses can be provided. In practice, less than 4.3 billion addresses are available due to address allocation reasons. In addition, IPv4 address resources are allocated unevenly. The USA uses almost half of the world's IP addresses, while Europe uses fewer IP addresses than the USA and the Asian-Pacific region uses even fewer. The development of mobile IP and broadband technologies requires more IP addresses. The shortage of IPv4 addresses directly restricts further development of the IP technology.

  There are several solutions to IPv4 address shortage. Classless Inter-domain Routing (CIDR) and Network Address Translation (NAT) are two major solutions. CIDR and NAT, however, have their disadvantages and outstanding problems. This also drives the need for and the development of IPv6.

- The backbone device needs to maintain a large number of routing entries.

  Due to allocation and planning problems in the early phase of IPv4 development, many discontinuous IPv4 addresses are allocated and as a result routes cannot be aggregated effectively. The increasingly large routing tables consume a lot of memory, leading to higher costs for equipment and lower forwarding efficiency. To tackle these issues, device manufacturers have to constantly upgrade their products in order to improve route addressing and forwarding performance.

- Address autoconfiguration and readdressing cannot be performed easily.

  An IPv4 address occupies only 32 bits and IP addresses are allocated unevenly. Consequently, IP addresses need to be reallocated during network expansion or network replanning. Address autoconfiguration and readdressing are required to simplify maintenance. With IPv4, however, such tasks cannot be performed easily.

- Security cannot be well guaranteed.

  With the development of the Internet, security problems become more serious. During the design of IPv4, security was not fully considered. Hence, the original framework cannot implement end-to-end security. IPv6 provides end-to-end security.

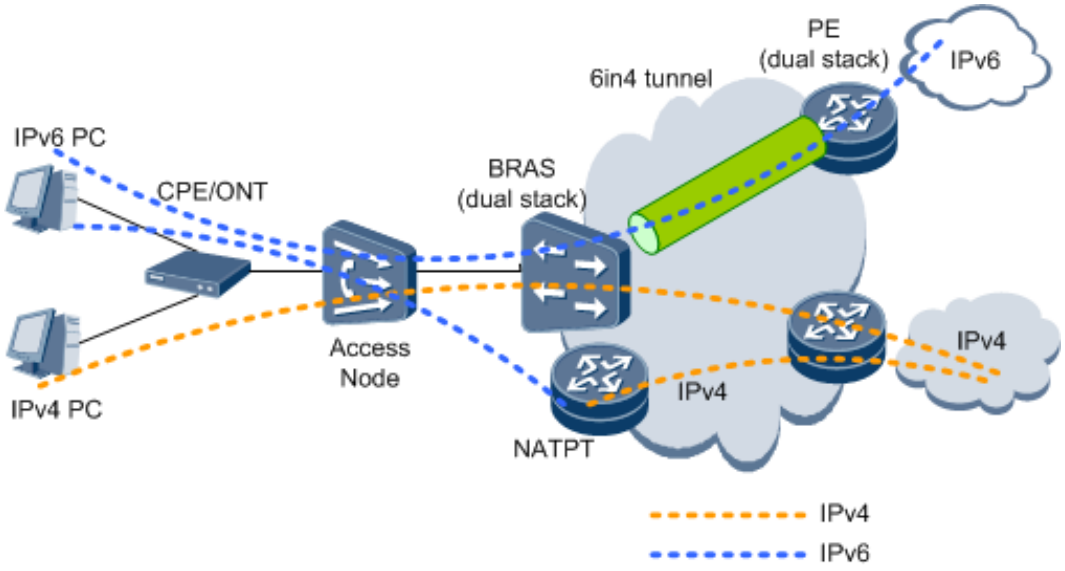
IPv6 solves the IP address shortage issue from the very outset. Moreover, IPv6 is easy to deploy, compatible with various applications, allows IPv4 networks to smoothly transit to IPv6 networks, and can coexist and interwork with IPv4. With so many advantages over IPv4, IPv6 is in rapid development.

11.2.2 IPv6 network deployment

Internet protocol version 6 (IPv6) is developed to tackle global IPv4 address shortage. At the early stage of IPv4-to-IPv6 transition, IPv4 networks have been in large-scale deployment whereas deployment of IPv6 networks is comparatively scarce, and both types of networks will coexist. **Figure 11-5** shows the IPv6 network deployment at the early stage of IPv4-to-IPv6 transition.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-5 Early-stage IPv6 network deployment



IPv4-based broadband remote access servers (BRASs) on the live network are progressively upgraded to support IPv4/IPv6 dual-stack. The dual-stack BRAS provides a 6in4 tunnel or a dedicated link to transmit IPv6 traffic to an IPv6 network.

The dual-stack BRAS provides the network address translation-protocol translation (NAT-PT) function to allow IPv6 users to access an IPv4 network.

The access network equipment is able to detect IPv6 packets, allocate IPv6 addresses (using DHCPv6), and support ACLv6.

11.2.3 IPv6 Principles

Basic functions of IPv6 include IPv6 neighbor discovery and IPv6 path maximum transmission unit (PMTU) discovery. Neighbor discovery and PMTU discovery are implemented through Internet Control Message Protocol for IPv6 (ICMPv6) messages.

11.2.3.1 IPv6 Highlights

- A 128-bit address structure, providing sufficient address space

  A major advantage of IPv6 is the almost infinite IP address space. IPv6 increases the size of an IP address from 32 bits to 128 bits, which is four times of that of IPv4. A 128-bit address structure is able to provide about $4{,}300{,}000{,}000^4$ addresses, meeting almost any address assignment requirements that can be predicted. (Theoretically, 4,300,000,000 IPv4 addresses can be provided; however, $4{,}300{,}000{,}000^4$IPv6 addresses can be provided.)

- Layered address structure

  The layered address structure realizes rapid route lookup, reduces the size of IPv6 routing tables with the aid of route aggregation, and thereby improves the forwarding efficiency of routers.

 Copyright © Huawei Technologies Co., Ltd.

- Address autoconfiguration

  IPv6 enables hosts to discover networks and obtain IPv6 addresses using address autoconfiguration, which greatly improves the network manageability. Using address autoconfiguration, user devices (such as mobile phones and wireless devices) support plug-and-play, without requiring manual configuration or using a private server (such as a DHCP server). IPv6 supports stateful address autoconfiguration and stateless address autoconfiguration.

  – In stateful address autoconfiguration, the host obtains the address and other configuration information from the server.

  – In stateless address autoconfiguration, the host automatically configures address information that contains the prefix and interface ID of the host as reported by the local router. If there is no router on the link, the host can automatically configure only a link-local address for interoperating with the local node.

- Source/Destination address selection

  To specify or plan source/destination addresses of the packets sent by the system, the network administrator can define a set of address selection rules. These rules form an address selection policy table. The policy table is similar to a routing table and employs the longest matching rule for prefix lookup. The address selection result is determined together by the source address and destination address.

  A source address is selected according to the following rules. Among the rules below, the rule with a smaller number has a higher priority. A candidate address is preferred if it:

  a. Is the same as the destination address.

  b. Has an appropriate effective scope.

  c. Is not a deprecated address.

  d. Is a home address.

  e. Is the address of an outgoing interface.

  f. Has the same *label* value as the destination address.

  g. Has the longest matching prefix.

  NOTE

  The candidate address can be a unicast address that is configured on the specified outgoing interface. If a source address that has the same *label* value as and is in the same address range with the destination address is not found on the outgoing interface, such a source address can be selected on another interface.

  A destination address is selected according to the following rules. Among the rules below, the rule with a smaller number has a higher priority. A candidate address is preferred if it:

  a. Is not an unusable address.

  b. Has an appropriate effective scope.

  c. Is not a deprecated address.

  d. Is a home address.

  e. Has the same *label* value as the source address.

 Copyright © Huawei Technologies Co., Ltd.

f. Has a larger *precedence* value.
g. Prefer native transport (6over4 or 6to4 tunnel is not required).
h. Has a smaller effective scope.
i. Has the longest matching prefix.
j. Otherwise, leave the order unchanged.

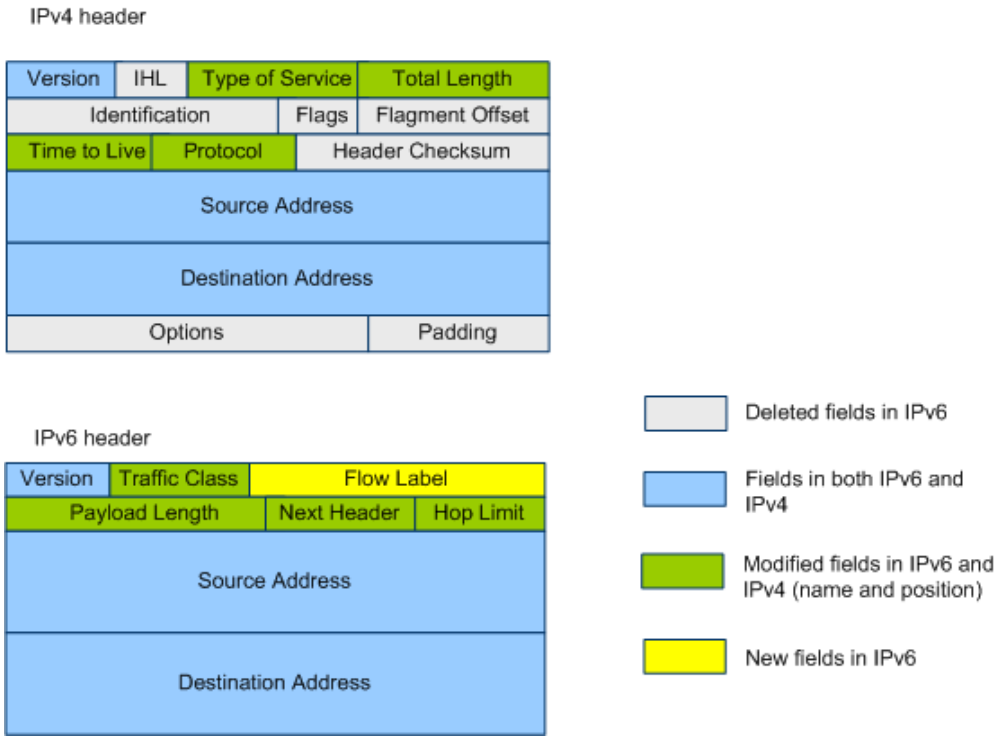
- Support for QoS

  New fields are added in the IPv6 header to define how to label and process flows. Flows are identified by the Flow Label field in the header. The Flow Label field allows routers to identify the packets of a certain flow and provide special processing for these packets.

- Flexible and simple extension headers

  Figure 11-6 shows the comparison between IPv6 and IPv4 header formats. Compared with an IPv4 header, an IPv6 header deletes the IHL, Identification, Flags, Fragment Offset, Header Checksum, Options, and Padding fields, and adds the Flow Label field, which helps improve the header processing efficiency. In addition, to better support different options, IPv6 introduces multiple extension headers. With these extension headers, it is unnecessary to modify the existing packet structure when new options are added, which greatly improves the flexibility of IPv6.

Figure 11-6 Comparison between IPv6 and IPv4 header formats



 Copyright © Huawei Technologies Co., Ltd.

11.2.3.2 IPv6 Addresses

Format of an IPv6 Address

A 128-bit IPv6 address has two formats:

- X:X:X:X:X:X:X:X
  - In this format, the 128 bits of an IPv6 address are divided into 8 groups. The 16 bits of each group are represented by 4 hexadecimal characters (0 to 9, and A to F). Groups are separated by colons. Every "X" represents the numerical value of a group of hexadecimal characters. The following is an example:

    2001:0DB8:130F:0000:0000:09C0:876A:130B

    For convenience, the zeros at the beginning of each group can be omitted. The preceding example, thus, can be written as 2001:DB8:130F:0:0:9C0:876A:130B.
  - Furthermore, the two or more consecutive zeros in the address can be replaced by "::", which reduces the written length of an IPv6 address. The preceding example can be further compressed as 2001:DB8:130F::9C0:876A:130B.

    An IPv6 address contains only one "::". Otherwise, a computer cannot determine the count of zeros when restoring the original 128-bit address from the compressed address.
- X:X:X:X:X:X:d.d.d.d

  Addresses in this format fall into the following two types:
  - IPv4-compatible IPv6 address: The format of an IPv4-compatible IPv6 address is 0:0:0:0:0:0:IPv4-address. The high-order 96 bits are all 0s, and the low-order 32 bits are an IPv4 address. This IPv4 address must be reachable in an IPv4 network, and cannot be a multicast, broadcast, loopback, or unspecified address (0.0.0.0).
  - IPv4-mapped IPv6 address: The format of an IPv4-mapped IPv6 address is 0:0:0:0:0:FFFF:IPv4-address. This type of IPv6 address is used to represent the addresses of IPv4 nodes.
  - An IPv4-compatible IPv6 address is used for the configuration of IPv6 over IPv4 tunnels.

    "X:X:X:X:X:X" represent the high-order 6 groups of numbers, and each "X" stands for 16 bits expressed in hexadecimal notation. "d.d.d.d" represent the low-order 4 groups of numbers, and each "d" stands for 8 bits expressed in decimal notation. "d.d.d.d" is a standard IPv4 address.

Structure of an IPv6 Address

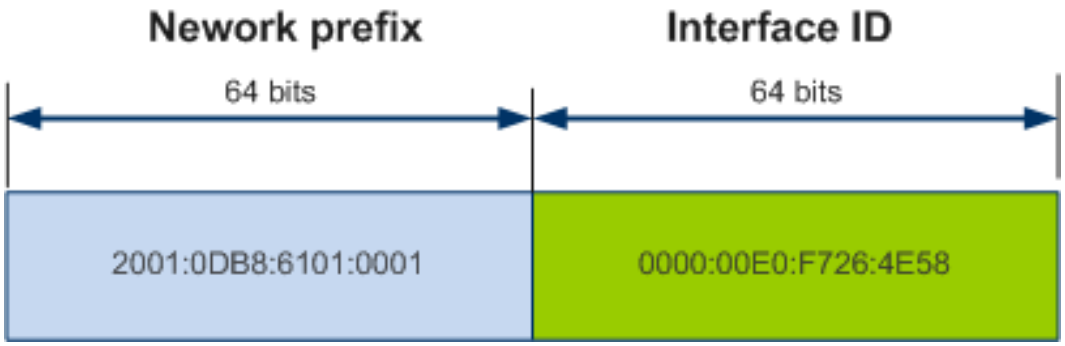
An IPv6 address can be divided into two parts:

- Network prefix: n bits, equaling the network ID in an IPv4 address.
- Interface identifier (ID): 128-n bits, equaling the host ID in an IPv4 address.

Figure 11-7 illustrates the structure of the address 2001:DB8:6101:1::E0:F726:4E58 /64.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-7 Structure of the address 2001:DB8:6101:1::E0:F726:4E58 /64



IPv6 Address Classification

IPv6 has the following types of addresses:

- Unicast address: uniquely identifies an interface and is similar to an IPv4 unicast address. The packets sent to a unicast address are transmitted to the unique interface identified by this address.

  Unicast addresses can be classified into the following types, as shown in **Table 11-1**.

  Table 11-1 Types of IPv6 unicast addresses

Address Type	Binary Prefix	IPv6 Prefix Identifier
Link-local unicast address	1111111010	FE80::/10
Loopback address	00...1 (128 bits)	::1/128
Unspecified address	00...0 (128 bits)	::/128
Global unicast address	Others	-

  The meanings of each type of address are described as follows:

  – Link-local IPv6 unicast address: It is used in the Neighbor Discovery (ND) protocol, and in the communication between nodes on the local link during the stateless address autoconfiguration. The packets carrying the link-local IPv6 unicast address as the source or destination address are forwarded only on the local link. A link-local IPv6 unicast address can be automatically configured on any interface by using the link-local prefix FE80::/10(1111 1110 10) and the interface identifier in the IEEE EUI-64 format (an EUI-64 identifier can be derived from an EUI-48 identifier).

  – Loopback address: It is 0:0:0:0:0:0:0:1 or ::1 and not assigned to any interface. Similar to the case of IPv4 loopback address 127.0.0.1, IPv6 packets carrying the IPv6 loopback address as the destination address are looped back to the sending node.

  – Unspecified address (::): It cannot be assigned to any node or function as a destination address. The unspecified address can be used in the Source Address field of the IPv6 packet sent by an initializing host that has not

 Copyright © Huawei Technologies Co., Ltd.

obtained its own address. During duplicate address detection (DAD), the Source Address field of a Neighbor Solicitation (NS) message is an unspecified address.

  - Global unicast address: It is equivalent to an IPv4 public network address. Global unicast addresses are used on the links that can be aggregated, and are provided to Internet service providers (ISPs). The structure of this type of address allows for route prefix aggregation to relieve the global routing entry resources limitation. A global unicast address consists of a 48-bit route prefix that is managed by the carrier, a 16-bit subnet ID that is managed by the local node, and a 64-bit interface ID. Unless otherwise specified, global unicast addresses include site-local unicast addresses.

- Anycast address: identifies a group of interfaces, which generally belong to different nodes. The packets carrying an anycast destination address are transmitted to the interface that is nearest to the source node in the interface group identified by the anycast address. The nearest interface refers to the interface with the smallest distance metric measured by the routing protocol.

  Application scenario: When a mobile host needs to communicate with the mobile agent on the home subnet, the mobile host uses the anycast address of the routing device in the subnet.

  Specifications of anycast addresses: Anycast addresses do not have independent address space. They can use the format of any unicast address. Thus, syntax is required to differentiate an anycast address from a unicast address.

- Multicast address: identifies a group of interfaces that belong to different nodes and is similar to an IPv4 multicast address. The packets carrying a multicast destination address are transmitted to all the interfaces identified by this multicast address.

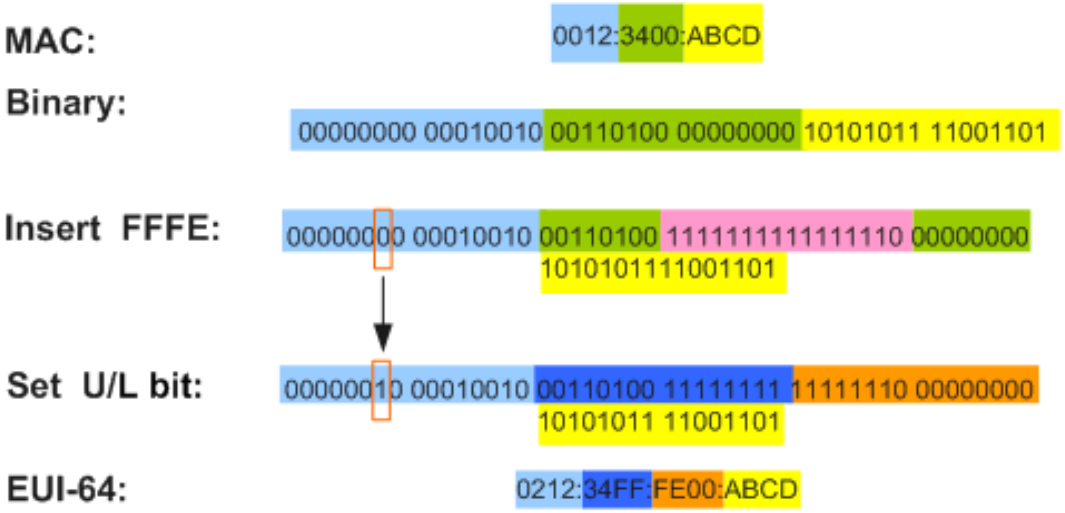
  IPv6 addresses do not include broadcast addresses. In IPv6, functions of broadcast addresses are provided by multicast addresses.

Interface ID in the IEEE EUI-64 Format

The 64-bit interface ID in an IPv6 address identifies a unique interface on a link. This address is derived from the link-layer address (such as a MAC address) of the interface. The interface ID of an IPv6 address contains 64 bits, while a MAC address contains 48 bits. By inserting a hexadecimal number FFFE (1111 1111 1111 1110) into the MAC address and then setting the U/L bit (the leftmost seventh bit) to 1, an interface ID in the EUI-64 format is generated. **Figure 11-8** shows conversion process.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-8 Converting a MAC address into an EUI-64 address

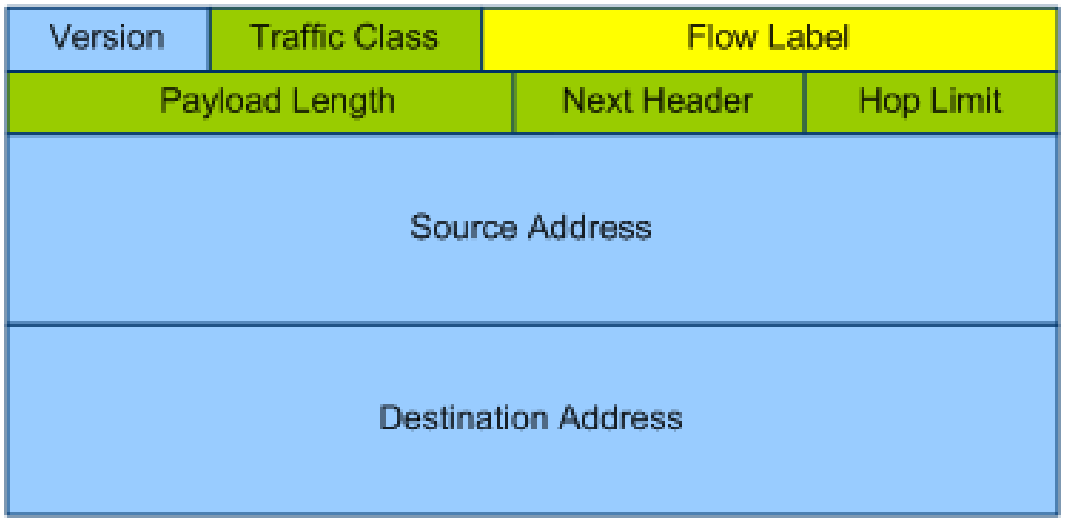


11.2.3.3 IPv6 Packet Format

Format of an IPv6 Header

Figure 11-9 shows the format of an IPv6 header.

Figure 11-9 Format of an IPv6 header



- Version:
  4 bits. The value of this field is 6, indicating an IPv6 packet.
- Traffic Class:
  8 bits. This field is similar to the TOS field in an IPv4 header.
- Flow Label:

 Copyright © Huawei Technologies Co., Ltd.

20 bits. This field is new in IPv6. Flow labels are used to label particular flows so as to differentiate packets at the network layer. Routers along a forwarding path differentiate and process packets according to the flow labels. Because the Flow Label field locates in the IPv6 header, forwarding routers and destination nodes do not need to differentiate packets according to the packet content.

- Payload Length:

  16 bits. This field indicates the length of the IPv6 payload in octets. The payload is the rest of the packet following the IPv6 header. (Note that any extension headers present are considered part of the payload.)

- Next Header:

  8 bits. This field identifies the type of header immediately following the current IPv6 header (which may be a header or an extension header). This field uses the same values as the IPv4 Protocol field. The Next Header field in the IPv6 header and the Next Header fields in the IPv6 extension headers form a chain. The chain mechanism helps improve the efficiency of extension header processing because the router processes only the option headers needed.

- Hop Limit:

  8 bits. This field is similar to the IPv4 TTL field. This field decreases by 1 by each node that forwards the packet. The packet is dropped if this field decreases to 0.

- Source Address:

  128 bits. This field indicates the source address of the packet.

- Destination Address:

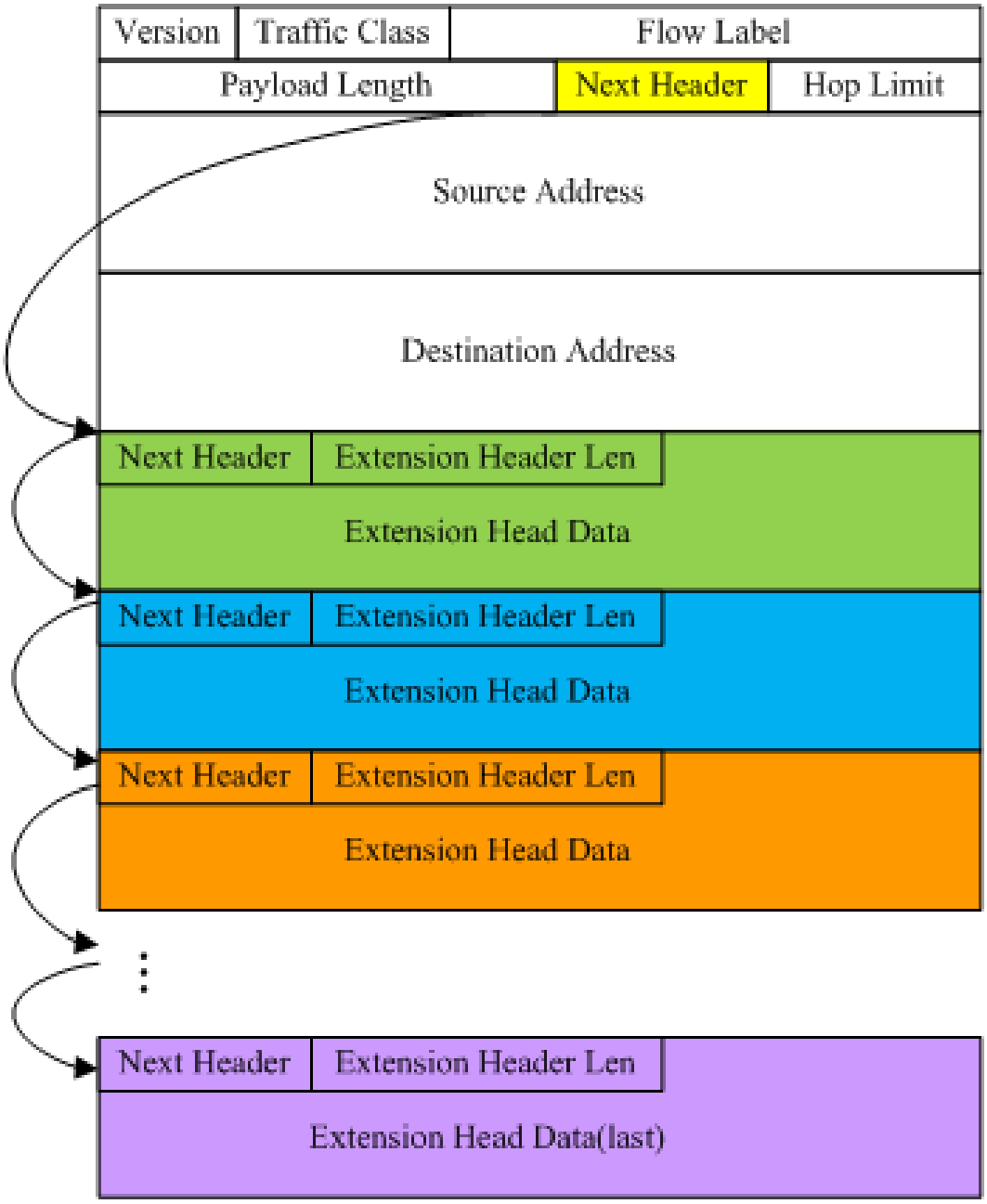
  128 bits. This field indicates the destination address of the packet.

Format of IPv6 Extension Headers

Figure 11-10 shows the format of IPv6 extension headers.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-10 Format of IPv6 extension headers



IPv6 option fields are supported through a chain of extension headers. An IPv6 packet can carry zero, one, or multiple extension headers.

IPv6 extension headers appear in the following order:

- Hop-by-Hop Options Header

  The value of this header is 0, which is defined in the IPv6 header. It is used for routing alarms and jumbo frames. This header is processed by every node along the packet forwarding path.
- Destination Options Header

 Copyright © Huawei Technologies Co., Ltd.

The value of this header is 60. This header may occur before the following two headers:

- Routing Header

  In such a case, the Destination Options header is processed by the destination node and the node specified in the Routing header.
- Upper-layer Header (located behind any ESP option)

  In such a case, the Destination Options header is processed only by the destination node.

- Routing Header

  The value of this header is 43. This header is used for source routing options and mobile IPv6.
- Fragment Header

  The value of this header is 44. This header is used for packet fragmentation when the packet sent by the source node is larger than the path maximum transmission unit (PMTU). PMTU is the MTU specified for the path from the source node to destination node.
- Authentication Header

  The value of this header is 51. This header is used for authentication and integrity checking of a packet. The definition of this header in IPv6 is the same as that in IPv4.
- ESP Header

  The value of this header is 50. This header is used for authentication, integrity checking, and encryption of a packet. The definition of this header in IPv6 is the same as that in IPv4.
- Upper-layer Header

  This header is an upper-layer protocol (such as TCP, UDP, or ICMP) header.

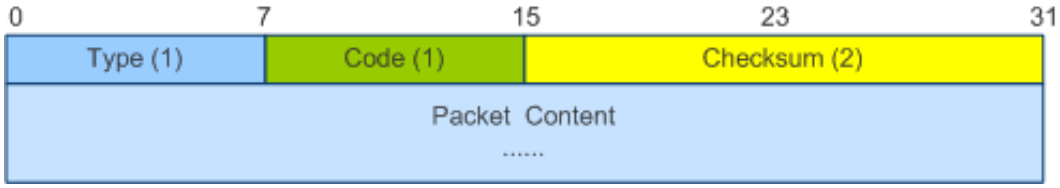
The Destination Options header occurs at most twice (once before the Routing header and once before the upper-layer header). The other extension headers occur at most once.

IPv6 nodes must be able to process the extension headers, regardless of the order and the number of occurrences they appear in the same packet, except for the Hop-by-Hop Options header that is restricted to appear immediately after an IPv6 header only. This requirement ensures interoperability between IPv6 nodes.

11.2.3.4 ICMPv6

As a basic protocol of IPv6, Internet Control Message Protocol for IPv6 (ICMPv6) is used by a device to generate two types of messages: error messages and informational messages. Using these messages, IPv6 nodes report errors and information generated during packet processing. **Figure 11-11** shows the format of an ICMPv6 message.

Figure 11-11 Format of an ICMPv6 message



 Copyright © Huawei Technologies Co., Ltd.

The meaning of each field in an ICMPv6 message is as follows:

- Type field: indicates the message type. The values from 0 to 127 indicate the error message type, and values from 128 to 255 indicate the informational message type.
- Code field: indicates the specific message type.
- Checksum field: indicates the checksum of an ICMPv6 message.

Classification of ICMPv6 Error Messages

- Destination Unreachable message

  When an IPv6 node is forwarding IPv6 packets and finds that the destination address of the packets is unreachable, it sends an ICMPv6 Destination Unreachable message to the source node of the packets. Specific causes for the error message are carried in the message. Destination Unreachable messages are classified into types that include the following:

  – No route to destination
  – Address unreachable
  – Port unreachable

- Packet Too Big message

  When an IPv6 node is forwarding IPv6 packets and finds that the size of the packets exceeds the path maximum transmission unit (PMTU) of the outbound interface, it sends an ICMPv6 Packet Too Big message to the source node of the packets. The PMTU of the outbound interface is carried in the message. PMTU discovery is implemented based on Datagram Too Big messages.

- Time Exceeded message

  During the reception and transmission of IPv6 packets, when a device receives a packet with the hop limit of 0 or when the device reduces the hop limit to 0, it sends an ICMPv6 Time Exceeded message to the source node of the packets. When reassembling fragmented packets, an ICMPv6 Time Exceeded message is also generated if the reassembly time exceeds the specified duration.

- Parameter Problem message

  When a destination node receives an IPv6 packet, it checks the validity of the packet. If the destination node detects any of the following errors, it sends an ICMPv6 Parameter Problem message to the source node of the packet:

  – A field in the IPv6 header or extension header is incorrect.
  – The Next Header field in the IPv6 header or extension header cannot be identified.
  – Unknown options exist in the extension header.

Classification of ICMPv6 Informational Messages

ICMPv6 informational messages are classified into Echo Request messages and Echo Reply messages. ICMPv6 informational messages can be used for network fault diagnosis, PMTU discovery, and neighbor discovery. During the interoperation

 Copyright © Huawei Technologies Co., Ltd.

check between two nodes, the node that receives an Echo Request message sends an Echo Reply message to the source node. In this manner, subsequent packets are received and transmitted between the two nodes.

NOTE

A large number of generated redirection ICMP packets will degrade the device performance. The device does not send redirection ICMP packets by default. The **icmpv6 redirect enable** command is used to enable the function of sending redirection ICMP packets.

11.2.3.5 PMTU

Problems Related to MTU

A path from a source address to a destination address may traverse interfaces that have different maximum transmission unit (MTU) values. The smallest MTU on this path is called the path MTU (PMTU).

- During transmission of IPv6 packets, the packets cannot be fragmented on the intermediate nodes. Therefore, it often happens that the packet length exceeds the PMTU. In such a case, the source node needs to retransmit the IPv6 packets, which reduces the transmission efficiency.
- If the minimum link IPv6 MTU (1280 bytes) is used as the maximum fragment length on the source node, in most cases, the IPv6 MTU of the path is greater than the IPv6 MTU of the minimum link, and the fragment sent by a node is far less than the IPv6 MTU of the path, which is a waste of network resources.

The PMTU discovery protocol is introduced to solve this problem.

Principle of PMTU Discovery

PMTU discovery describes a method of dynamically discovering the PMTU for a path. When an IPv6 node sends a large amount of data to another node, the data is transmitted by means of a sequence of IPv6 fragments. When these fragments are of the maximum length allowed in successful transmission between the source node and destination node, the fragment length is considered optimal and called PMTU.

A source node assumes that a PMTU of a path is the known IPv6 MTU of the first hop on the path. If the packet sent from this path is too large to be forwarded along the path, the intermediate node discards this packet and returns an ICMPv6 Packet Too Big message to the source node. The source node then sets the PMTU of the path according to the IPv6 MTU specified in the message.

When the PMTU learned by the node is smaller than or equal to the actual PMTU, the PMTU discovery process ends. Before the PMTU discovery process end, ICMPv6 Packet Too Big messages may be repeatedly sent and received because smaller IPv6 MTUs may be found further down the path.

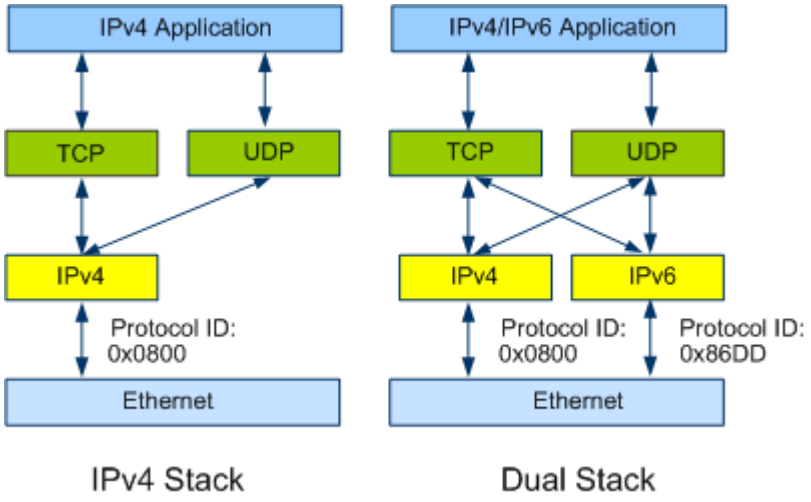
11.2.3.6 Dual Protocol Stacks

For an IPv6 node, the most effective way of being compatible with IPv4 is to retain a complete IPv4 protocol stack on the node. Such a node is called a dual-stack

 Copyright © Huawei Technologies Co., Ltd.

node. **Figure 11-12** shows the structures of a single protocol stack and dual protocol stacks.

Figure 11-12 Structures of a single protocol stack and dual protocol stacks in an Ethernet



Dual protocol stacks have the following advantages:

- Multiple link-layer protocols support dual protocol stacks.

  Multiple link-layer protocols, such as Ethernet, support dual protocol stacks. In **Figure 11-12**, the link-layer protocol is Ethernet. In an Ethernet frame, if the Protocol ID field is 0x0800, it indicates that the network layer receives IPv4 packets; if the field is 0x86DD, it indicates that the network layer receives IPv6 packets.
- Multiple applications support dual protocol stacks.

  The upper layer applications, such as the DNS, can use TCP or UDP as the transmission layer protocol, and prefers the IPv6 protocol stack rather than the IPv4 protocol stack as the network-layer protocol.

11.2.3.7 TCP6

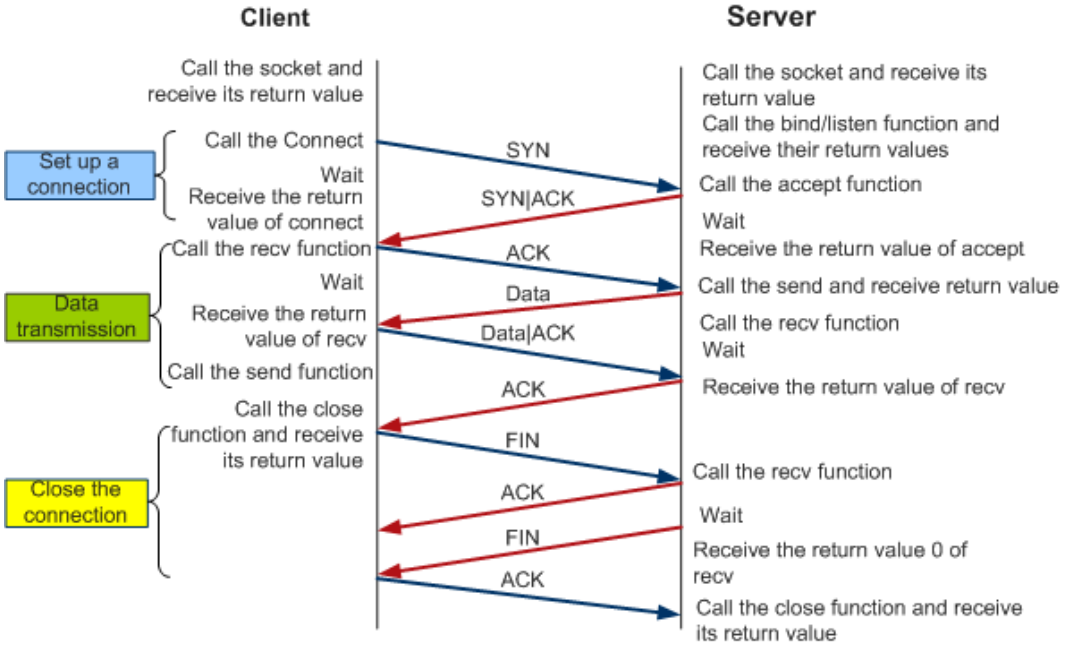
Transmission Control Protocol Version 6 (TCP6) provides a mechanism to establish virtual circuits between processes of two endpoints. A TCP6 virtual circuit is similar to the full-duplex circuit that transmits data between systems. TCP6 is called a reliable protocol because it provides reliable data transmission between processes. TCP6 also provides a mechanism to optimize the transmission performance according to the network status. When all the data can be received and acknowledged, the transmission rate increases gradually.

TCP6 is generally used in interactive applications, such as the web application. TCP6 establishes virtual circuits by using the three-way handshake mechanism, and all the virtual circuits are deleted using the four-way handshake mechanism. TCP6 connections provide a variety of checksums and reliability-ensuring functions, but increase the cost. As a result, TCP6 has lower efficiency than User Datagram Protocol Version 6 (UDP6).

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-13 shows the establishment and removal of a TCP6 connection.

Figure 11-13 Establishment and removal of a TCP6 connection



11.2.3.8 UDP6

User Datagram Protocol Version 6 (UDP6) is a computer communication protocol used to exchange packets on a network. UDP6 has the following characteristics:

- UDP uses only source and destination information and runs mainly in a simple request/response structure.
- UDP is unreliable. No control mechanism is provided to ascertain whether UDP6 datagrams have reached their destinations.
- UDP is connectionless. No virtual circuits are required for data transmission between hosts.

The connectionless feature of UDP6 enables UDP6 to send data to broadcast addresses. This is different from TCP6, which requires specific source and destination addresses.

11.2.3.9 RawIP6

RawIP6 is an implementation in which only a limited number of fields in the IPv6 header are filled with values, and RawIP6 allows applications to provide their own IPv6 headers.

RawIP6 is similar to UDP6 in the following aspects:

- RawIP6 is unreliable. No control mechanism is provided to ascertain whether RawIP6 datagrams have reached their destinations.
- RawIP6 is connectionless. No virtual circuits are required for data transmission between hosts.

 Copyright © Huawei Technologies Co., Ltd.

Unlike UDP6, RawIP6 allows applications to directly operate the IP layer through sockets. Therefore, RawIP6 is convenient for the applications that need to interact directly with the lower layer.

11.2.3.10 IPv6 Neighbor Discovery

The Neighbor Discovery (ND) protocol defines a set of messages and processes for determining the relationship between neighboring nodes. The IPv6 ND protocol supports the Address Resolution Protocol (ARP) messages, and the Router Discovery and Redirect messages of the Internet Control Message Protocol (ICMP) of IPv4, and also provides other functions.

After an IPv6 address is configured on a node, the node first verifies that the address is available and does not conflict with other addresses. If the node is a host, the router needs to notify the host of a better first-hop address to reach a particular destination. If the node is a router, this node needs to advertise its address, address prefix, and other configuration information, which provides guidance for the host to configure parameters. During IPv6 packet forwarding, a node needs to verify the link-layer address and reachability of its neighboring node. IPv6 ND defines five types of ICMPv6 packets:

- Router Solicitation (RS): Sent by an enabled host to a router. The router then responds with a Router Advertisement (RA).
- RA: Advertised by a router periodically, which contains information such as the prefix and some flags.
- Neighbor Solicitation (NS): Sent by an IPv6 node to determine the link-layer address of a neighbor, to check whether a neighbor is reachable, or for duplicate address detection (DAD).
- Neighbor Advertisement (NA): A response to an NS. An IPv6 node may also send unsolicited NAs to announce a link-layer address change.
- Redirect: Used by a router to inform hosts of a better first hop for a specific destination when the router finds that the incoming interface and outgoing interface of the packets are the same.

The ND protocol for IPv6 has the following functions.

DAD

In an IPv6 network, a link-local address (LLA) is used for the communication between neighboring nodes on the same link, for example, the communication between hosts on a single link where no routers exist. If an LLA is available, it is automatically used for neighbor discovery. Duplicate address detection (DAD) is a detection mechanism used for determining whether an LLA address is available. The process is as follows:

1. When an IPv6 address is configured on a node, the node sends an NS to its neighboring node to check whether the IP address is already used.
2. When receiving the NS, the neighboring node checks whether it has the same IPv6 address. If yes, the neighboring node responds with an NA carrying the IPv6 address information.
3. The node, upon receiving the NA, considers that the IPv6 address configured is already used by its neighboring node. If the node does not receive any

 Copyright © Huawei Technologies Co., Ltd.

response packet from its neighboring node, the IPv6 address configured is available.

DAD Proxy

A DAD proxy is used to prevent LLA conflicts on a Layer 3 interface. The DAD proxy feature resolves the issue of LLA conflicts between isolated ports on the same Layer 3 interface.

- When a user performs DAD, the MA5800 checks whether any LLA conflict occurs in the Layer 3 interface VLAN based on the information about the packet sent by the user.
  - If no conflict occurs, the MA5800 forwards the packet.
  - If a conflict occurs, the MA5800 responds to this conflict and drops the packet.
- The MA5800 obtains the LLA lease time from Router Advertisement (RA) and DHCPv6 packets, and the buffered LLA entries age when the lease time expires.

Neighbor Discovery

ND for IPv6, similar to ARP in IPv4, is used to parse the addresses of neighbors and detect whether neighbors are reachable using NSs and NAs.

To obtain the link-layer address of another node in the same local link, a node (source node) sends an NS with its ICMPv6 type as 135. This packet is similar to an ARP request packet in IPv4; but unlike the ARP request packet using a broadcast address, the NS uses a multicast address. A node with the last 24 bits of its address the same as the multicast address will receive this NS, which reduces the possibility of broadcast storm. The node receiving the NS (destination node) fills in its link-layer address in the response packet.

An NS can also be used to check whether a neighboring node is reachable when the link-layer address of the neighboring node is known. An IPv6 NA message is sent in response to an IPv6 NS message. The destination node, upon receiving an NS, responds with an NA with its ICMPv6 type as 136 over the local link. The source node then is able to communicate with the destination node after receiving the NA. A node may also send unsolicited NAs to announce a link-layer address change on the local link.

Router Discovery

Router Discovery is used to locate neighboring routers as well as learn prefixes and configuration parameters related to stateless address autoconfiguration (SLAAC). Router Discovery in IPv6 is implemented using the following two mechanisms:

- Router Solicitation (RS) message

  When unicast addresses are not configured on a host (for example, the system just starts up), the host sends an RS. An RS facilitates the host autoconfiguration, without having to wait for the RA sent by IPv6 routers. The IPv6 RS message is an ICMPv6 message of Type 133.
- Router Advertisement (RA) message

 Copyright © Huawei Technologies Co., Ltd.

Each RA-enabled IPv6 router periodically sends RAs. After receiving an RS from an IPv6 node on the local link, an IPv6 router also responds with an RA. An IPv6 router sends an RA to the multicast addresses (FF02::1) of all nodes or to the IPv6 unicast address of the node sending an RS. An RA is an ICMPv6 of type 134, including the following content:

- Whether to use address autoconfiguration
- Supported autoconfiguration type (stateless or stateful)
- One or more local link prefixes (nodes on the local link can use these prefixes to perform address autoconfiguration)
- Lifetime of local link prefixes advertised
- Whether the router sending an RA can serve as a default router. If yes, the information also contains the lifetime of the default router in units of seconds.
- Other host-related information, such as the hop limit and the MTU used for the host to send packets

An IPv6 node on the local link receives an RA and updates information such as the default router, prefix list, and other information from this RA.

Address Autoconfiguration

RAs and per-prefix flags enable routers to inform hosts how to perform address autoconfiguration. For example, routers can specify whether hosts use stateful (DHCPv6) or stateless (SLAAC) address configuration.

When using the SLAAC protocol, a host uses the prefix information and local interface ID obtained from the RA received to automatically generate an IPv6 address. Also, the host can set the default router according to the default router information in the RA.

Redirect

Redirect messages are sent by routers to inform a host of a better first-hop IPv6 address for a specific destination. Like IPv4, IPv6 Redirect messages are sent only for redirecting packets to a better router. Nodes receiving these Redirect messages will send packets to this better router. Routers send Redirect messages only for unicast flows, and Redirect messages are only sent to and processed by those nodes (hosts) triggering redirect.

Default Router Priority and Routing Information

An RA defines two fields: default router priority and routing information, which helps hosts select a better forwarding router for packets.

When the link where a host resides has multiple routers, the host needs to select the forwarding router according to the packet's destination address. In such cases, routers advertise default router priority and specific routing information to hosts, improving hosts' capability of selecting better forwarding routers according to different destination addresses.

A host, upon receiving an RA that contains the routing information, will update its routing list. Before sending packets to other devices, the host selects a better route according to its routing list.

 Copyright © Huawei Technologies Co., Ltd.

A host updates its default router list after receiving an RA that contains default router priority information. Before sending packets to other devices, the host queries its default router list and selects the highest-priority router for sending packets if no route is available. If the highest-priority router is faulty, the host will choose the second-highest-priority router and so on.

ND Proxy

NOTE

This function is supported in V100R020C00 and later versions.

User-side packets are sent to only the upper-layer device and they cannot be forwarded between user sides. After multicast domains are isolated, user networks are more secured. However, user terminals require interconnection, that is, different VLANs or some users isolated in the same VLAN need to communicate with each other.

Both IPv4 and IPv6 networks have such a requirement. It is similar to ARP proxy for IPv4 networks. ND proxy is an expansion of ARP proxy for IPv6 networks, which is used to resolve the IPv6 network interconnection issue. To configure it,

- Enable the ND proxy on a super VLAN interface, and then the super VLAN interface acts as the proxy of all the sub VLANs in the super VLAN.
- To enable the ND proxy for the NEs in a sub VLAN, the ND proxy of the sub VLAN must be enabled.

11.2.4 Configuring Basic IPv6 Information

This topic describes the IPv6 features supported by the MA5800. The basic IPv6 configuration includes configuration of the IPv6 address, IPv6 neighbor discovery, path maximum transmission unit (PTMU), and transmission control protocol 6 (TCP6).

Context

The following table lists the IPv6 features supported by the MA5800.

Table 11-2 IPv6 features supported by the MA5800

Feature	Sub-feature	Configuration Process or Command
IPv6 address management and assignment	Static configuration of IPv6 global unicast addresses and IPv6 link-local addresses	**Configuring an IPv6 Address for an Interface**
	Automatic configuration of IPv6 link-local addresses	**ipv6 address auto link-local**
	DHCPv6, DHCPv6 L2/L3 Relay	**11.7.4 Configuring DHCPv4 Relay**

 Copyright © Huawei Technologies Co., Ltd.

Feature	Sub-feature	Configuration Process or Command
	Management information base (MIB) for IPv6 address management	-
IPv6 stack and IPv6 host function	IPv6/IPv4 dual-stack to ensure compatibility of IPv6 and IPv4	-
	Basic IPv6 protocols, including ICMPv6, TCP6, UDP6, and RawIP6	-
	IPv6 Neighbor Discovery (ND) protocol and static configuration of IPv6 neighbors	**Configuring IPv6 Neighbor Discovery**
	IPv6 PMTU	**Configuring PMTU**
	IPv6 ping and tracert	• **ping ipv6** • **tracert ipv6**
	IPv6 statistics query and clearance	• **display ipv6 statistics** • **reset ipv6 statistics**
IPv6 route	IPv6 static routes	**11.12.2.5 Configuration Example of the IPv6 Static Route**
	BGP4+	**11.12.8.4 Configuration Example of BGP4+**
IPv6 QoS and security	IPv6 ACL	**Configuring Traffic Management Based on ACL Rules**
	Anti-MAC spoofing	**Configuring MAC Anti-Spoofing**
	Anti-IPv6 spoofing	**Configuring IPv6/ICMPv6 Address Attack Prevention**
	Anti-denial of service (DoS) attack	**Configuring Anti-DoS Protection**
DAD Proxy and Proxy advertisement for NS	DAD Proxy	**ipv6 dad proxy**
	Proxy advertisement for neighbor solicitation (NS) on the network side	-

 Copyright © Huawei Technologies Co., Ltd.

Feature	Sub-feature	Configuration Process or Command
IPv6 Layer 2 Transparent Transmission	Differentiation of service virtual ports based on the IPv6 over Ethernet (IPv6oE) type (0x86DD) and defining of VLANs for service virtual ports	**service-port**
	Transparent transmission of IPv6 over PPPoE packets	-
	VLAN-based transparent transmission of IPv6 packets	-

NOTE

In this manual:

- For the IPv6 features that are different from IPv4 features, configuration procedures and examples are provided for both IPv6 and IPv4 features.
- For the IPv6 features that are similar with IPv4 features, configuration procedures and examples are not provided for IPv6 features because they are the same as IPv4 features. To configure these IPv6 features, use IPv6 commands and follow the procedures of IPv4 features.

11.2.4.1 Configuring an IPv6 Address for an Interface

The MA5800 can communicate with other IPv6 equipment only after its interface is configured with an IPv6 address. Before an IPv6 global unicast address or IPv6 link-local address is configured on an interface, the IPv6 packet forwarding function must be enabled on the device.

Context

Each interface can be configured with a maximum of 10 IPv6 global unicast addresses but only one IPv6 link-local address.

- An IPv6 global unicast address is equivalent to an IPv4 public address. It is used for forwarding data across the public network and is necessary for the communication between users. An EUI-64 address has the same function as an IPv6 global unicast address. The difference is that only the network bits need to be specified for the EUI-64 address and the host bits are transformed from the media access control (MAC) addresses of the interface, while a complete 128-bit address needs to be specified for the IPv6 global unicast address.
- The IPv6 link-local address is used in neighbor discovery (ND), and for the communication between nodes on the local link in the stateless address autoconfiguration (SLAAC) process. The packets using the link-local address as the source or destination address are not forwarded to other links.

  The link-local address can be automatically generated or manually configured. It is recommended to automatically generate a link-local address because the

 Copyright © Huawei Technologies Co., Ltd.

link-local address is used to implement communication requirements of protocol and is not directly related to the communication between users.

NOTE

In the SLAAC process, a host uses the prefix information and local interface ID obtained from the received router advertisement (RA) to automatically generate an IPv6 address, rather than using the stateful address autoconfiguration mechanism (DHCPv6).

The MA5800 supports IPv6 address configuration on the VLANIF interface, METH interface, and loopback interface. This topic uses the VLANIF interface as an example.

Procedure

Step 1 Enable IPv6 packet forwarding capability.

Enabling the IPv6 function on the device and the interface is a prerequisite for configuring IPv6 features. To enable a device to forward IPv6 packets, you must enable the IPv6 capability in both the global config mode and the interface mode.

By default, the IPv6 function is disabled on the device and interface.

1. In global config mode, run the **interface vlanif** command to enter the VLANIF mode.
2. In VLANIF mode, run the **ipv6 enable** command to enable the IPv6 function on the interface.

NOTE

Before configuring other IPv6 features on an interface, you must enable the IPv6 function in interface mode.

Step 2 Configure an IPv6 global unicast address on the interface.

In VLANIF mode, run the **ipv6 address** or **ipv6 address eui-64** command to configure an IPv6 global unicast address on the interface.

To implement successful communication, you can configure both the EUI-64 address and the IPv6 global unicast address, or configure only one of them. The IPv6 addresses configured on one interface cannot be in the same network segment.

Step 3 Configure an IPv6 link-local address on the interface.

In VLANIF mode, use either of the following methods to configure the IPv6 link-local address:

- Run the **ipv6 address auto link-local** command to automatically generate a local-link address on the interface.

  After this command is executed, the deletion of the global unicast address does not affect local link communication. If the device only needs to communicate with another device that is directly connected to the device, using the link-local address saves IPv6 global unicast address resources.
- Run the **ipv6 address link-local** command to manually configure a link-local address on the interface. The prefix of the IPv6 address configured by running this command must be FE80::/10.

 Copyright © Huawei Technologies Co., Ltd.

If an automatically allocated link-local address exists on the interface, the link-local address will be overwritten by the new link-local address after this command is executed. If possible, avoid changing the link-local address.

If you have not run either of the preceding commands to configure a link-local address for an interface, a link-local address will be automatically allocated to the interface after an IPv6 global unicast address is configured on the interface.

Step 4 Query the address configuration information about an IPv6 interface.

- Run the **display ipv6 interface** command to query the IPv6 interface information.
- Run the **display ipv6 statistics** command to query the IPv6 packet statistics.

----End

Example

To create VLAN 10, set the IPv6 address of VLANIF interface 10 to 2001:db8:1::1/64, and set the generation mode of the link-local address to the automatic generation mode, run the following commands:

```
huawei(config)#vlan 10
huawei(config)#interface vlanif 10
huawei(config-if-vlanif10)#ipv6 enable
huawei(config-if-vlanif10)#ipv6 address 2001:db8:1::1/64
huawei(config-if-vlanif10)#ipv6 address auto link-local
```

11.2.4.2 Configuring PMTU

By setting the path maximum transmission unit (PMTU), the device can select a proper maximum transmission unit (MTU) for packet transmission. In this manner, packets do not have to be fragmented during transmission. Employing PMTU reduces the load on intermediate devices, improves network resources utilization, and achieves optimal throughput on the network.

Context

Dynamic PMTU is enabled on a device by default, ensuring that the smallest MTU value is used on all interfaces along the path from the source to the destination node. You can also configure static PMTU to specify the maximum length of a packet that can be forwarded from the source to the destination node. Configuring static PMTU protects devices on a network from jumbo packets. Static PMTU has a higher priority over dynamic PMTU. When both static PMTU and dynamic PMTU are configured, only static PMTU takes effect.

Procedure

Step 1 Configure static PMTU.

In global config mode, run the **ipv6 pathmtu** command to configure a static PMTU value for a path destined for a specified IPv6 address. By default, the PMTU of the path destined for an IPv6 address is 1500 bytes.

The static PMTU value should be smaller than or equal to the MTU value of every interface on the same path. If the static PMTU value is larger than the MTU value of the interfaces, the system segments packets according to the MTU value. By

 Copyright © Huawei Technologies Co., Ltd.

manually configuring the static PMTU value based on the smallest MTU value of the path over which packets are transmitted, you can achieve a higher packet transmission rate.

Step 2 Configure the aging time of dynamic PMTU entries.

Run the **ipv6 pathmtu age** command to configure the aging time of dynamic PMTU entries. By default, the aging time of dynamic PMTU entries is 10 minutes.

The **ipv6 pathmtu age** command is used to modify the lifetime of dynamic PMTU entries in the buffer. The aging time is invalid on static PMTU entries, because static PMTU entries do not age.

Step 3 Query the PMTU information.

- Run the **display ipv6 pathmtu** command to query the PMTU information.
- Run the **display ipv6 interface** command to query the current MTU value of an IPv6 interface.

----End

Example

To set the PMTU for the IPv6 address 2001:db8:: to 1300 bytes, and to set the aging time of the dynamic PMTU entries to 40 minutes, run the following commands.

```
huawei(config)#ipv6 pathmtu 2001:db8:: 1300
huawei(config)#ipv6 pathmtu age 40
```

11.2.4.3 Configuring IPv6 Neighbor Discovery

The IPv6 Neighbor Discovery (ND) protocol provides a set of packets and processes for establishing the relationship between neighboring nodes. The IPv6 ND protocol supports the Address Resolution Protocol (ARP) messages, and the Router Discovery and Redirect messages of the Internet Control Message Protocol (ICMP) of IPv4. In addition, ND supports neighbor reachability detection.

Prerequisite

The IPv6 address has been configured. For details on the configuration method, see **Configuring an IPv6 Address for an Interface**.

Context

Most of the ND configurations are implemented based on the interface. The MA5800 supports ND configurations on VLAN Layer 3 interfaces.

Procedure

- Configure the static IPv6 neighbor.

  By configuring a static neighbor, the device can obtain the mapping of the IPv6 address and MAC address of the neighbor. The statically configured neighbor entries will overwrite the dynamically learned neighbor entries and will not age.

 Copyright © Huawei Technologies Co., Ltd.

a. In global config mode, run the **interface vlanif** command to enter the VLANIF mode.

b. In VLANIF mode, run the **ipv6 neighbor** command to configure a static IPv6 neighbor.

- Configure the parameters of the Router Advertisement (RA) message.

  The device periodically sends RA messages that contain information such as address prefixes, hop limit value, neighbor reachable time, and message lifetime. The IPv6 node on the local link receives the RA messages and updates its information, such as the IPv6 prefix list and other configuration data, according to the RA messages.

  a. In global config mode, run the **interface vlanif** command to enter the VLANIF mode.

  b. Run the **undo ipv6 nd ra halt** command to enable RA message advertising.

     - When a device is connected to an IPv6 node, the RA message advertising function needs to be enabled so that the device periodically sends RA messages to the IPv6 node.
     - When a device is not connected to an IPv6 node, the RA message advertising function does not need to be enabled. By default, this function is disabled.

     By default, RA message advertising is disabled on the device.

  c. (Optional) Configure parameters carried in an RA message.

     Perform the following operations as needed:

     - Run the **ipv6 nd ra** command to configure the interval for advertising RA messages.

       By default, the maximum interval is 600 seconds and the minimum interval is 200 seconds. The maximum interval cannot be shorter than the minimum interval. When the maximum interval is shorter than 9 seconds, the minimum interval is adjusted to the same value as the maximum interval.

     - Run the **ipv6 nd ra router-lifetime** command to configure the RA message lifetime.

       By default, the lifetime is 1800 seconds. The lifetime of the message advertised by the device must be longer than or equal to the interval at which the device advertises RA messages.

     - Run the **ipv6 nd ra prefix** command to configure the address prefixes to be advertised in RA messages.

       By default, RA messages contain only the address prefixes specified through the **ipv6 address** command. Run the **ipv6 nd ra prefix** command when you need the device to advertise only the specified prefixes.

     - Set the stateful autoconfiguration flag bit carried in an RA message.

 Copyright © Huawei Technologies Co., Ltd.

- Run the **ipv6 nd autoconfig managed-address-flag** command to set the flag bit carried in an RA message for stateful address autoconfiguration.

  After this flag bit is set, the host uses a stateful address autoconfiguration protocol; otherwise, the host uses a stateless address autoconfiguration protocol.

- Run the **ipv6 nd autoconfig other-flag** command to set the flag bit for stateful autoconfiguration of other information.

  After this flag bit is set, the host uses a stateful autoconfiguration protocol for configuring information other than the address.

- Run the **ipv6 nd ns retrans-timer** command to configure the interval for detecting neighbor reachability, that is, the neighbor solicitation (NS) message retransmission timer.

  Frequently sending NS packets helps to determine the neighbor reachability but also affects the device performance. Therefore, you are not advised to set the interval to a small value. The default interval, 1000 milliseconds, is recommended.

- Run the **ipv6 nd nud reachable-time** command to configure the neighbor reachable time specified in an RA message.

  The device detects neighbor reachability by using the neighbor unreachability detection (NUD) mechanism. The neighbor reachable time configured through the **ipv6 nd nud reachable-time** command is the interval for running a NUD detection. A smaller neighbor reachable time set on a device means that the device can probe the neighbor reachability more quickly but, meanwhile, more network bandwidth and CPU resources will be consumed. Therefore, you are not advised to set the neighbor reachable time to a short interval. The default interval, 30000 milliseconds, is recommended.

- Configure the hop limit value for the router.

  Run the **quit** command to quit the VLANIF mode. In global config mode, run the **ipv6 nd hop-limit** command to configure the hop limit value for the router, that is, the maximum number of hops for the IPv6 unicast packets initiated by the router. The hop limit value for the router is the same as the hop limit value specified in the RA message. The default value is 64.

- Configure duplicate address detection (DAD).

  DAD is used to check whether an IPv6 address is available. When a node is configured with an IPv6 address, it immediately sends an NS message to check whether this address is already used by other neighboring nodes.

  a. In global config mode, run the **interface vlanif** command to enter the VLANIF mode.
  b. Run the **ipv6 nd dad attempts** command to configure the number of DAD attempts, that is, number of attempts to send NS messages. The default value is 1.

 Copyright © Huawei Technologies Co., Ltd.

c. Run the **ipv6 nd ns retrans-timer** command to configure the interval of DAD, that is, the timer for retransmitting NS messages. The default interval, 1000 milliseconds, is recommended.

- Query the IPv6 neighbor information.
  - Run the **display ipv6 neighbors** command to query the IPv6 neighbor information.
  - Run the **display ipv6 interface** command to query the IPv6 interface information.

----End

Example

To configure the function of automatically generating a link-local unicast address on VLANIF interface 10, set the prefix to be advertised by the local unicast address of site EUI-64 and by the RA message to 2001:db8:3::2/64, set both the valid lifetime and the preferred lifetime of the prefix to 1000s, and enable RA message advertising (so that the host can automatically obtain the address prefix in the RA message), run the following commands:

```
huawei(config)#vlan 10
huawei(config)#interface vlanif 10
huawei(config-if-vlanif10)#ipv6 enable
huawei(config-if-vlanif10)#ipv6 address auto link-local
huawei(config-if-vlanif10)#ipv6 address 2001:db8:3::2/64 eui-64
huawei(config-if-vlanif10)#ipv6 nd ra prefix 2001:db8:3::2/64 1000 1000  { <cr>|no-autoconfig<K>|off-link<K> }:  Command:  ipv6 nd ra prefix 2001:db8:3::2/64 1000 1000
huawei(config-if-vlanif10)#undo ipv6 nd ra halt
```

11.2.5 IPv6 Reference Standards and Protocols

The following table lists the reference standards and protocols of the IPv6 feature.

Standard/ Protocol	Description
RFC1887	An Architecture for IPv6 Unicast Address Allocation
RFC1981	Path MTU Discovery for IP version 6
RFC2375	IPv6 Multicast Address Assignments
RFC2460	Version 6 of the Internet Protocol (IPv6), also sometimes referred to as IP Next Generation or IPng.
RFC2461/ RFC4861	Neighbor Discovery for IP Version 6 (IPv6)
RFC2462/ RFC4862	IPv6 Stateless Address Auto configuration
RFC2463/ RFC4443	Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification
RFC2464	Transmission of IPv6 Packets over Ethernet Networks

 Copyright © Huawei Technologies Co., Ltd.

Standard/ Protocol	Description
RFC2466	Management Information Base for IP Version 6 ICMPv6 Group
RFC2711	IPv6 Router Alert Option
RFC2893	Transition Mechanisms for IPv6 Hosts and Routers
RFC3315	DHCPv6 snooping
RFC3484	Default Address Selection for Internet Protocol Version 6 (IPv6)Section 2.1
RFC3493	Basic Socket Interface Extensions for IPv6
RFC3633	IPv6 Prefix Options for Dynamic Host Configuration Protocol(DHCP) version 6
RFC3736	Stateless Dynamic Host Configuration Protocol (DHCP) Service for IPv6
RFC3849	IPv6 Address Prefix Reserved for Documentation
RFC4001	Textual Conventions for Internet Network Addresses
RFC4007	IPv6 Scoped Address Architecture
RFC4191	Default Router Preferences and More-Specific Routes
RFC4193	Unique Local IPv6 Unicast Addresses
RFC4213	Basic Transition Mechanisms for IPv6 Hosts and Routers
RFC4214	Intra-Site Automatic Tunnel Addressing Protocol(ISATAP)
RFC4429	Duplicate Address Detection
RFC4282	A Model of IPv6/IPv4 Dual Stack Internet Access Service
RFC2373/ RFC3513/ RFC4291	Internet Protocol Version 6 (IPv6) Addressing Architecture
RFC4862/ RFC5006	Router Advertisement (RA) filtering
RFC6221	DHCPv6 LDRA

11.3 ARP

 Copyright © Huawei Technologies Co., Ltd.

11.3.1 Introduction

Definition

The Address Resolution Protocol (ARP) is an Internet protocol used to map IP addresses to MAC addresses.

Purpose

To communicate with each other, the sending device must know the network-layer IP address of the receiving device. IP datagrams, however, must be encapsulated with MAC addresses before they can be transmitted over the physical network. Therefore, a resolution from the known IP address to a MAC address is required.

ARP was developed to enable communication between devices.

Function Overview

Table 11-3 lists ARP features.

Table 11-3 ARP features

Feature	Description	Usage Scenario
Dynamic ARP	Devices dynamically learn and update the mapping between IP addresses and MAC addresses by exchanging ARP messages.	Real-time communication is a priority, or network resources are insufficient.
Static ARP	The mapping between IP addresses and MAC addresses is manually created and cannot be dynamically modified.	Communication security is a priority, and network resources are sufficient.
Gratuitous ARP	A device broadcasts gratuitous ARP packets that carry the local IP address as both the source and destination IP addresses to notify the other devices on the same network segment of its address information.	Gratuitous ARP is used to check whether the local IP address conflicts with that of another device, to notify other devices on the same network segment of the new MAC address after the local network interface card is replaced, or to notify master/slave switchovers in a Virtual Router Redundancy Protocol (VRRP) backup group.

 Copyright © Huawei Technologies Co., Ltd.

Benefits

ARP ensures Ethernet communication by mapping IP addresses at the network layer to MAC addresses at the link layer on Ethernet networks.

11.3.2 Principles

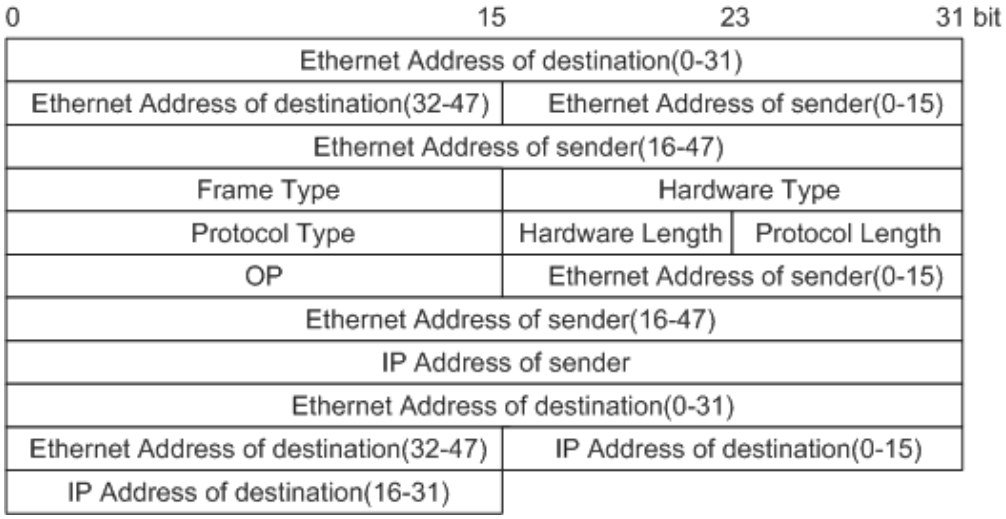
11.3.2.1 Basic Principles

Related Concepts

- **Address Resolution Protocol (ARP) messages**

  An ARP message can be an ARP request or reply message. **Figure 11-14** shows the ARP message format.

  Figure 11-14 ARP message format



  NOTE

  The **Ethernet Address of destination** field contains a total of 48 bits. **Ethernet Address of destination (0-31)** indicates the first 32 bits of the **Ethernet Address of destination** field, and **Ethernet Address of destination (32-47)** indicates the last 16 bits of the **Ethernet Address of destination** field.

  An ARP message consists of 42 bytes. The first 14 bytes indicate the Ethernet frame header, and the last 28 bytes are the ARP request or reply message content. **Table 11-4** describes the fields in an ARP message.

 Copyright © Huawei Technologies Co., Ltd.

Table 11-4 Description of fields in an ARP message

Field	Length	Description
Ethernet Address of destination	48 bits	Ethernet destination MAC address in the Ethernet frame header. This field in an ARP Request message is the broadcast MAC address, with a value of 0xFF-FF-FF-FF-FF-FF.
Ethernet Address of sender	48 bits	Ethernet source MAC address in the Ethernet frame header.
Frame Type	16 bits	Data type. For an ARP request or reply message, the value of this field is 0x0806.
Hardware Type	16 bits	Hardware address type. For an Ethernet network, the value of this field is 1.
Protocol Type	16 bits	Type of the protocol address to be mapped by the sending device. For an IP address, the value of this field is 0x0800.
Hardware Length	8 bits	Hardware address length. For an ARP request or reply message, the value of this field is 6.
Protocol Length	8 bits	Protocol address length. For an ARP request or reply message, the value of this field is 4.
OP	16 bits	Operation type. The values are as follows: ● 1: ARP request ● 2: ARP reply ● 3: RARP request ● 4: RARP reply
Ethernet Address of sender	48 bits	Source MAC address. The value of this field is the same as the Ethernet source MAC address in the Ethernet frame header.
IP Address of sender	32 bits	Source IP address.
Ethernet Address of destination	48 bits	Destination MAC address. The value of this field in an ARP Request message is 0x00-00-00-00-00-00.
IP Address of destination	32 bits	Destination IP address.

- **ARP table**

  An ARP table contains the latest mapping between IP addresses and MAC addresses. If a host always broadcasts an ARP Request message for a MAC address before it sends an IP datagram, network communication traffic will greatly increase. Furthermore, all other hosts on the network have to receive and process the ARP Request messages, which lowers network efficiency. To

 Copyright © Huawei Technologies Co., Ltd.

solve this problem, an ARP table is maintained on each host to ensure efficient ARP operations. The mapping between an IP address and a MAC address is called an ARP entry.

ARP entries can be classified as dynamic or static.

- Dynamic ARP entries are automatically generated and maintained by using ARP messages. Dynamic ARP entries can be aged and overwritten by static ARP entries.
- Static ARP entries are manually configured and maintained by a network administrator. Static ARP entries can neither be aged nor be overwritten by dynamic ARP entries.

Before sending IP datagrams, a host searches the ARP table for the MAC address corresponding to the destination IP address.

- If the ARP table contains the corresponding MAC address, the host directly sends the IP datagrams to the MAC address instead of sending an ARP Request message.
- If the ARP table does not contain the corresponding MAC address, the host broadcasts an ARP Request message to request the MAC address of the destination host.

- **Reverse Address Resolution Protocol (RARP)**

  If only the MAC address of a host is available, the host can send and receive RARP messages to obtain its IP address.

  To do so, the network administrator must establish the mapping between MAC addresses and IP addresses on a gateway. When a new host must be configured, its RARP client requests the host's IP address from the RARP server on the gateway.

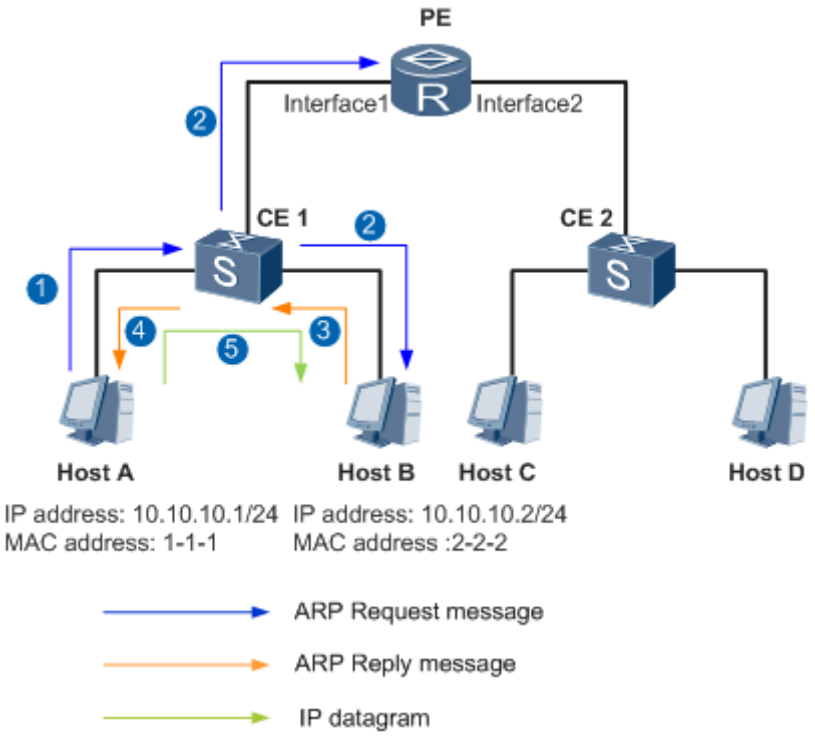
Implementation

- **ARP implementation within a network segment**

  Figure 11-15 illustrates how ARP is implemented within a network segment, by using IP datagram transmission from Host A to Host B as an example.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-15 ARP implementation between Host A and Host B on the same network segment



a. Host A searches its ARP table and does not find the mapping between the IP address and MAC address of Host B. Host A then sends an ARP Request message for the MAC address of Host B. In this ARP Request message, the source IP address and source MAC address are respectively the IP address and MAC address of Host A, the destination IP address and destination MAC address are respectively the IP address of Host B and 00-00-00-00-00-00, and the Ethernet source MAC address and Ethernet destination MAC address are respectively the MAC address of Host A and the broadcast MAC address.

b. After CE1 receives the ARP Request message, CE1 broadcasts it on the network segment.

c. After Host B receives the ARP Request message, Host B adds the MAC address of Host A to its ARP table and sends an ARP Reply message to Host A. In this ARP Reply message, the source IP and MAC addresses are respectively the IP and MAC addresses of Host B, the destination IP and MAC addresses are respectively the IP and MAC addresses of Host A, and the Ethernet source and destination MAC addresses are respectively the MAC addresses of Host B and Host A.

   NOTE

   The PE also receives the ARP Request message but discards it because the destination IP address in the ARP Request message is not its own IP address.

d. CE1 receives the ARP Reply message and forwards it to Host A.

e. After Host A receives the ARP Reply message, host A adds the MAC address of Host B to its ARP table and sends the IP datagrams to Host B.

 Copyright © Huawei Technologies Co., Ltd.

- **ARP implementation between different network segments**

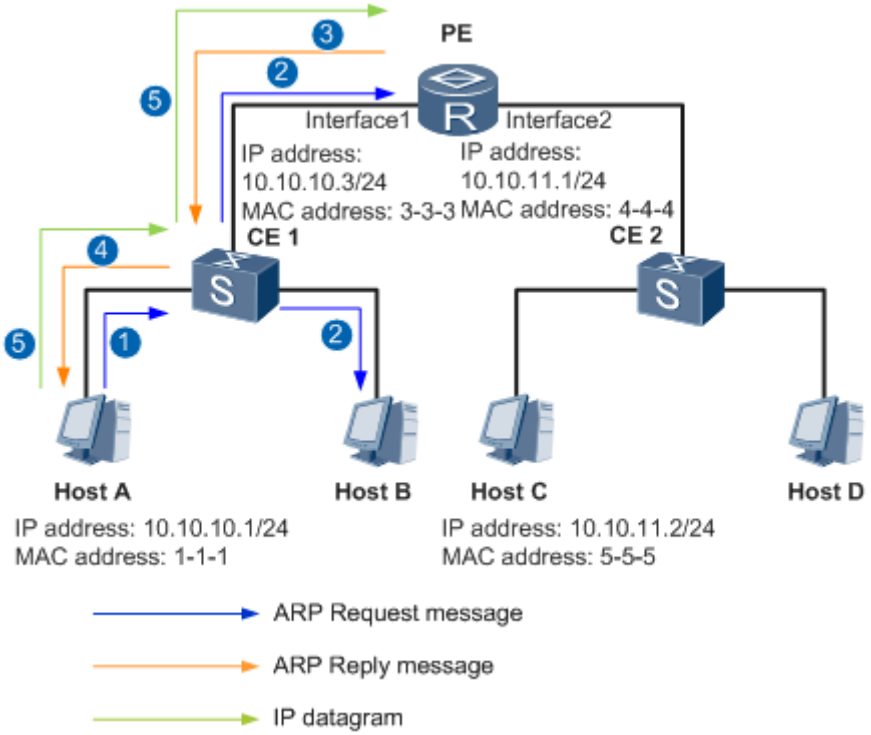
  NOTE

  ARP messages are Layer 2 messages. Therefore, ARP is applicable only to devices on the same network segment. If two hosts on different network segments need to communicate, the source host sends IP datagrams to the default gateway, which in turns forwards the IP datagrams to the destination host. ARP implementation between different network segments involves separate ARP implementation within network segments. In this manner, hosts on different network segments can communicate.

  The following examples show how ARP is implemented between different network segments, by using IP datagram transmission from Host A to Host C as an example.

  Figure 11-16 illustrates how ARP is implemented between Host A and the PE on the same network segment.

  Figure 11-16 ARP implementation between Host A and the PE



  a. Host A searches its ARP table and does not find the mapping between the IP address and MAC address of Interface 1 on the default gateway PE that connects to Host C. Host A then sends an ARP Request message for the MAC address of the PE's Interface. In this ARP Request message, the source IP and MAC addresses are respectively the IP and MAC addresses of Host A, the destination IP and MAC addresses are respectively the IP address of the PE's Interface 1 and 00-00-00-00-00-00, and the Ethernet source and destination MAC addresses are respectively the MAC address of Host A and the broadcast MAC address.

  b. After CE1 receives the ARP Request message, CE1 broadcasts it on the network segment.

 Copyright © Huawei Technologies Co., Ltd.

c. After the PE receives the ARP Request message, the PE adds the MAC address of Host A to its ARP table and sends an ARP Reply message to Host A. In this ARP Reply message, the source IP and MAC addresses are respectively the IP and MAC addresses of the PE's Interface 1, the destination IP and MAC addresses are respectively the IP and MAC addresses of Host A, and the Ethernet source and destination MAC addresses are respectively the MAC address of Interface 1 on PE and the MAC address of Host A.

NOTE

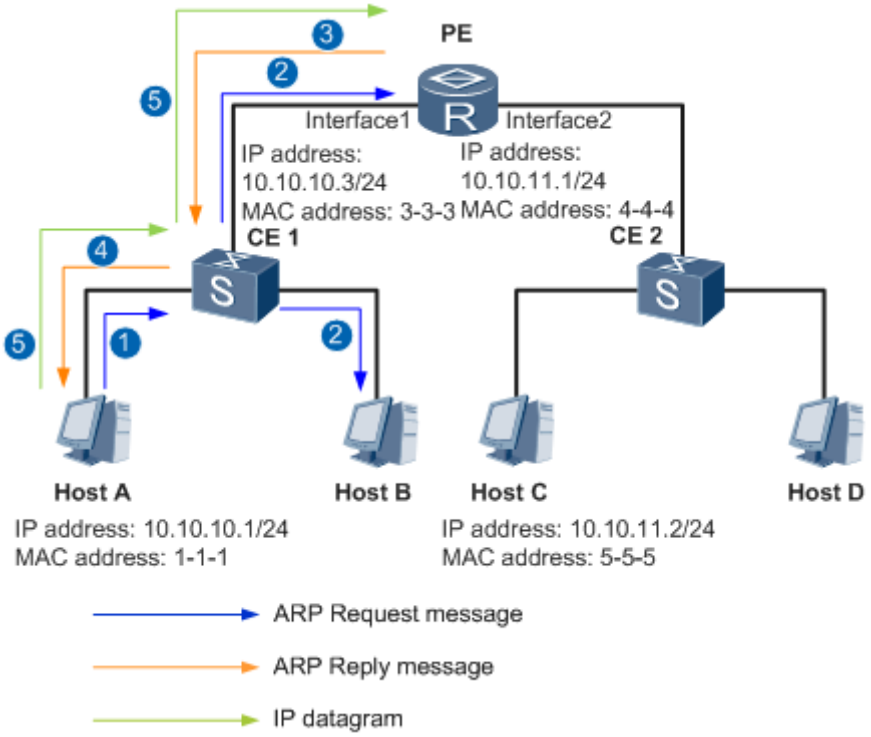
Host B also receives the ARP Request message but discards it because the destination IP address in the ARP Request message is not its own IP address.

d. CE1 receives the ARP Reply message and forwards it to Host A.

e. After Host A receives the ARP Reply message, host A adds the MAC address of the PE's Interface 1 to its ARP table and sends the IP datagrams to the PE.

Figure 11-17 illustrates ARP implementation between the PE and Host C on the same network segment.

Figure 11-17 ARP implementation between the PE and Host C



The PE searches its routing table and sends the IP datagrams from Interface 1 to Interface 2.

a. The PE searches its ARP table and does not find the mapping between the IP address and MAC address of Host C. Then, the PE sends an ARP Request message for the MAC address of Host C. In this ARP Request message, the source IP and MAC addresses are respectively the IP and

 Copyright © Huawei Technologies Co., Ltd.

MAC addresses of the PE's Interface 2, the destination IP and MAC addresses are respectively the Host C's IP address and 00-00-00-00-00-00, and the Ethernet source and destination MAC address are respectively the MAC address of Interface 2 on PE and the broadcast MAC address.

b. After CE2 receives the ARP Request message, CE2 broadcasts it on the network segment.

c. After Host C receives the ARP Request message, Host C adds the MAC address of the PE's Interface 2 to its ARP table and sends an ARP Reply message to the PE. In this ARP Reply message, the source IP and MAC addresses are respectively the IP and MAC addresses of Host C, the destination IP and MAC addresses are respectively the IP and MAC addresses of the PE's Interface 2, and the Ethernet source and destination MAC addresses are respectively the MAC address of Host C and the MAC address of Interface 2 on PE.

NOTE

Host D also receives the ARP Request message but discards it because the destination IP address in the ARP Request message is not the IP address of itself.

d. CE2 receives the ARP Reply message and forwards it to the PE.

e. After the PE receives the ARP Reply message, the PE adds the MAC address of Host C to its ARP table and sends the IP datagrams to Host C.

So far, the IP datagram transmission from Host A to Host C is complete.

NOTE

1. ARP Request messages are broadcast, whereas ARP Reply messages are unicast.
2. In ARP implementation, the switches CE1 and CE2 transparently forward IP datagrams and do not modify them.

11.3.2.2 Dynamic ARP

Definition

The Dynamic Address Resolution Protocol (ARP) indicates that devices dynamically learn and update the mapping between IP addresses and MAC addresses by using ARP messages. You do not need to manually configure the mapping.

Related Concepts

- Dynamic ARP involves the dynamic ARP aging mechanism.

  The dynamic ARP aging mechanism enables an ARP entry that is not used over a specified period to be automatically deleted. This mechanism helps reduce storage space of ARP tables and speed up ARP table queries.

  Table 11-5 describes concepts related to the dynamic ARP aging mechanism.

 Copyright © Huawei Technologies Co., Ltd.

Table 11-5 Concepts related to the dynamic ARP aging mechanism

Concept	Description	Usage Scenario
Aging probe mode	Before a dynamic ARP entry on a device is aged, the device sends ARP aging probe messages to the other devices on the same network segment. An ARP aging probe message can be a unicast or broadcast message. By default, a device broadcasts ARP aging probe messages.	• If the IP address of the peer device remains unchanged but its MAC address changes frequently, it is recommended that you configure ARP aging probe messages to be broadcast. • If the MAC address of the peer device remains unchanged, network bandwidth resources are insufficient, and the aging time of ARP entries is set to a small value, it is recommended that you configure ARP aging probe messages to be unicast.
Aging time	A dynamic ARP entry has a life cycle. If a dynamic ARP entry is not updated before its life cycle ends, this dynamic ARP entry is deleted from the ARP table. The life cycle is called aging time.	Two interconnected devices can learn the mapping between their IP and MAC addresses using ARP and can save the mapping in their ARP tables. Then, the two devices can communicate by using the ARP entries. When the peer device becomes faulty, or the network adapter of the peer device is replaced but the local device does not receive any status change information about the peer device, the local device continues sending IP datagrams to the peer device. As a result, network traffic is interrupted because the ARP table of the local device is not promptly updated. To reduce the risk of network traffic interruption, an aging timer can be set for each ARP entry. After the aging timer of a dynamic ARP entry expires, the entry is automatically deleted.

 Copyright © Huawei Technologies Co., Ltd.

Concept	Description	Usage Scenario
Number of aging probe attempts	Before a dynamic ARP entry is aged, the local device sends ARP aging probe messages to the peer device. If the local device does not receive an ARP Reply message after the number of aging probe attempts reaches a specified number, the dynamic ARP entry is deleted.	The ARP aging timer can help reduce the risk of network traffic interruptions that occur because an ARP table is not updated quickly enough, but cannot eliminate problems due to delays. Specifically, if the dynamic ARP entry aging time is *N* seconds, the local device can detect the status change of the peer device after *N* seconds. During the *N* seconds, the ARP table of the local device is not updated. If the number of aging probe attempts is specified, the local device can obtain the status change information about the peer device and update its ARP table.

Implementation

Dynamic ARP entries can be created, updated, and aged.

- Creating and updating dynamic ARP entries

  If a device receives an ARP message that meets either of the following conditions, the device automatically creates or updates an ARP entry:

  - The source IP address of the ARP message is on the same network segment as the IP address of the inbound interface. The destination IP address of the ARP message is the IP address of the inbound interface.
  - The source IP address of the ARP message is on the same network segment as the IP address of the inbound interface. The destination IP address of the ARP message is the virtual IP address of the Virtual Router Redundancy Protocol (VRRP) backup group configured on the interface on the device.

- Aging dynamic ARP entries

  After the aging timer of a dynamic ARP entry on a device expires, the device sends ARP aging probe messages to the peer device. If the device does not receive an ARP Reply message after the number of aging probe attempts reaches a specified number, the dynamic ARP entry is aged.

NOTE

The MA5800 supports semi-static ARP for a specified IP address or interface. Semi-static ARP entries are dynamically learned and can be updated but cannot be aged. After the semi-static ARP attribute is deleted, the generated ARP entry can be aged.

Usage Scenario

Dynamic ARP is applicable to a network with a complex topology, insufficient bandwidth resources, and a high requirement for real-time communication.

 Copyright © Huawei Technologies Co., Ltd.

Benefits

Dynamic ARP entries are dynamically created and updated using ARP messages. They do not need to be manually maintained, greatly reducing maintenance workload.

11.3.2.3 Static ARP

Definition

The Static Address Resolution Protocol (ARP) indicates that the mapping between IP addresses and MAC addresses is manually created by a network administrator.

Principles

The main difference between static ARP and dynamic ARP is how ARP entries are generated and maintained. Dynamic ARP entries are automatically generated and maintained using ARP packets. Static ARP entries are manually configured and maintained by network administrators. The advantages and disadvantages of dynamic ARP and static ARP are as follows:

- Dynamic ARP
  - Advantages: Dynamic ARP entries free network administrators from manual configuration and maintenance. Especially when the network device is faulty or the network adapter of the host is frequently changed, the ARP entries can be dynamically updated in real time, which greatly reduces the maintenance workload of the network administrator.
  - Disadvantages:
    - Dynamic ARP entries can be aged or overwritten by new dynamic ARP entries, which cannot ensure the stability and security of network communication.
    - The execution of dynamic ARP consumes certain network resources. Therefore, dynamic ARP does not apply to networks with insufficient bandwidth resources and may affect user services.
- Static ARP
  - Advantages:
    - Static ARP entries are not aged or overwritten by dynamic ARP entries, ensuring network stability.
    - With static ARP, IP addresses and MAC addresses are bound to prevent network attackers from tampering with ARP entries using ARP packets, ensuring network communication security.
    - Static ARP eliminates the need of dynamic ARP execution and reduces network resource consumption.
  - Disadvantages: Static ARP entries need to be manually configured by network administrators, causing heavy maintenance workload when it comes to frequent changes of the network structure.

To ensure the stability and security of network communication, deploy static ARP based on actual requirements and network resources.

 Copyright © Huawei Technologies Co., Ltd.

Static ARP can implement the following functions:

- Binds IP addresses to the MAC address of a specified gateway so that IP datagrams destined to these IP addresses must be forwarded by this gateway.
- Binds the destination IP addresses of IP datagrams sent by a specified host to a nonexistent MAC address, helping filter out unwanted IP datagrams.

Related Concepts

Static ARP entries are classified as short or long entries.

- Short static ARP entries

  Short static ARP entries contain only IP and MAC addresses. A device still has to send ARP Request messages. If the source IP and MAC addresses of the received reply messages are the same as the configured IP and MAC addresses, the device adds the interface that receives the ARP Reply messages to the static ARP entries. The device can use this interface to forward subsequent messages directly. Short static ARP entries cannot be directly used to forward messages

  Configure short static ARP entries to enable the device and a specified host to communicate using fixed IP and MAC addresses.

  NOTE

  In Network Load Balancing (NLB) scenarios, you must configure both MAC entries with multiple outbound interfaces and short static ARP entries for the gateway. These MAC entries and short static ARP entries must have the same MAC address. In this scenario, short static ARP entries are also called ARP entries with multiple outbound interfaces, and cannot be updated manually.

- Long static ARP entries

  Long static ARP entries contain IP and MAC addresses as well as the VLAN and outbound interface through which devices send packets. Long static ARP entries are directly used to forward messages.

  Configure long static ARP entries to further specify the interface through which only the host can communicate with the device in a VLAN.

Usage Scenario

Static ARP applies to the following scenarios:

- Networks with a simple topology and high stability
- Networks on which information security is of high priority, such as a government or military network
- Short static ARP entries mainly apply to scenarios in which network administrators want to bind hosts' IP and MAC addresses but hosts' access interfaces can change.

Benefits

Static ARP ensures communication security. If a static ARP entry is configured on a device, the device can communicate with the peer device using only the specified MAC address. Network attackers cannot modify the mapping between the IP address and MAC address by using ARP messages, ensuring normal communication between the two devices.

 Copyright © Huawei Technologies Co., Ltd.

11.3.2.4 Gratuitous ARP

Principles

To ensure the stability and reliability of network communication, a device can broadcast gratuitous Address Resolution Protocol (ARP) messages to notify the other devices on the same network segment of its address information in the following scenarios:

- You need to check whether the IP address of a device conflicts with the IP address of another device on the same network segment. The IP address of each device must be unique to ensure the stability of network communication.
- After the MAC address of a host changes after its network adapter is replaced, the host must quickly notify other devices on the same network segment of the MAC address change before the ARP entry is aged. This ensures the reliability of network communication.
- When a master/slave switchover occurs in a Virtual Router Redundancy Protocol (VRRP) backup group, the new master router needs to notify other devices on the same network segment of its status change.

Related Concepts

Gratuitous ARP involves gratuitous ARP messages.

A gratuitous ARP message is a special ARP message that carries the sending device's IP address as both the source and destination IP addresses.

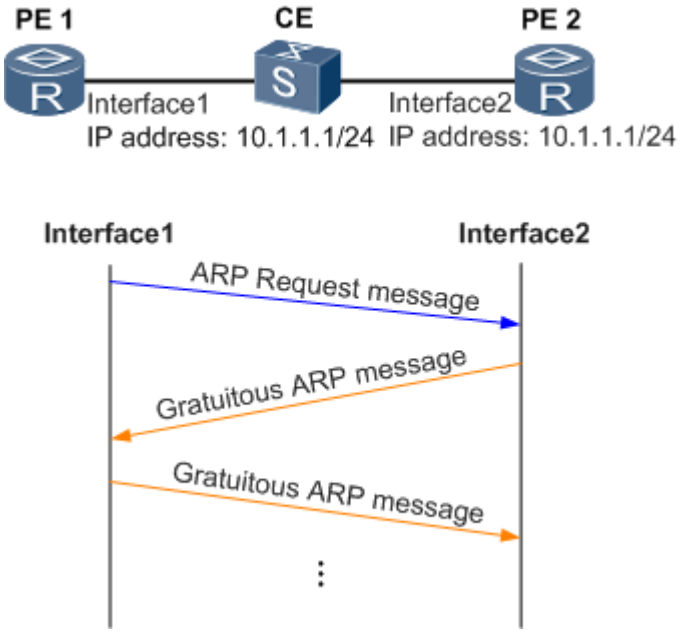
Implementation

- If a device finds that the source IP address in a received gratuitous ARP message is the same as its own IP address, the device sends a gratuitous ARP message to notify the sending device of the address conflict.
- If a device finds that the source IP address in a received gratuitous ARP message is different from its own IP address, the device updates the corresponding ARP entry with the sending device's IP address and MAC address carried in the gratuitous ARP message.

Figure 11-18 illustrates how gratuitous ARP is implemented.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-18 Gratuitous ARP implementation



As shown in **Figure 11-18**, the IP address of Interface 1 on PE1 is 10.1.1.1, and the IP address of Interface 2 on PE2 is 10.1.1.1.

1. Interface 1 broadcasts an ARP Request message. Interface 2 receives the ARP Request message and finds that the source IP address in the message conflicts with its own IP address. Then, Interface 2 performs the following operations:
   a. Interface 2 sends a gratuitous ARP message to notify Interface 1 of its IP address.
   b. A conflict node is generated on the conflict link of Interface 2. Interface 2 then sends gratuitous ARP messages to Interface 1 at an interval of 5 seconds.
2. Interface 1 receives the gratuitous ARP messages from Interface 2 and finds that the source IP address in the message conflicts with its own IP address. Interface 1 then performs the following operations:
   a. Interface 1 sends a gratuitous ARP message to notify Interface 2 of its IP address.
   b. A conflict node is generated on the conflict link of Interface 1. Interface 1 then sends gratuitous ARP messages to Interface 2 at an interval of 5 seconds.

Interface 1 and Interface 2 send gratuitous ARP messages to each other at an interval of 5 seconds until the address conflict is rectified.

If one interface does not receive a gratuitous ARP message from the other interface within 8 seconds, the interface considers the address conflict rectified. The interface deletes the conflict node on its conflict link and stops sending gratuitous ARP messages to the other interface.

 Copyright © Huawei Technologies Co., Ltd.

Functions

Gratuitous ARP implements the following functions:

- To check for IP address conflict on the network, send a gratuitous ARP message from a device. If the device receives a gratuitous ARP message from another device, the IP addresses of the two devices conflict.
- When the MAC address of a host changes after its network adapter is replaced, the host sends a gratuitous ARP message to notify other devices of the MAC address change before the ARP entry is aged. This ensures the reliability of network communication. After receiving the gratuitous ARP message, other devices maintain the corresponding ARP entry in their ARP tables based on the address information carried in the message.
- When a master/slave switchover occurs in the VRRP backup group, the new master router sends a gratuitous ARP message to notify other devices on the network of its status change.

Benefits

Gratuitous ARP reveals address conflict on a network so that ARP tables of devices can be quickly updated. This feature ensures the stability and reliability of network communication.

11.3.2.5 IP Address Conflict Detection

Background

The occurrence of an IP address conflict causes route flapping and traffic interruptions, affecting user services. IP address conflicts are often caused by incorrect networking or configurations. Users expect that devices can automatically detect IP address conflicts on a network and immediately notify users of conflict reasons, so that they can rapidly resolve such conflicts and minimize impact on services.

IP address conflict detection can help users quickly locate and modify the conflicted IP addresses and instruct users to properly configure and manage the IP addresses of devices on a network.

Implementation

IP address conflict detection can be classified into active and passive detection, and their differences are as follows:

- Active detection

  When the physical status of an interface on a device changes to Up, the device actively sends gratuitous ARP packets to detect possible IP address conflicts. For the detailed detection procedure, see **11.3.2.4 Gratuitous ARP**.

- Passive detection

  When a device receives ARP packets that are not gratuitous ARP packets, it checks the IP addresses carried by the ARP packets. The device concludes that IP address conflicts exist on the network if any of the following conditions are met:

 Copyright © Huawei Technologies Co., Ltd.

- The source IP address in an ARP packet is the same as the IP address of the inbound interface that receives the ARP packet, but the source MAC address in the ARP packet is different than the MAC address of the inbound interface.
- The source IP address in an ARP packet is the same as the IP address in an existing ARP entry, but the source MAC address is different than the MAC address in the ARP entry.
- The source IP address in an ARP packet is different than the CE IP address configured on the inbound interface that connects to the CE, or the source MAC address is different than the CE MAC address configured on the inbound interface that connects to the CE.

Usage Scenario

IP address conflict detection is applicable to Ethernet LANs.

Benefits

IP address conflict detection helps users quickly locate and modify IP address conflicts to ensure stability and security of user services.

11.3.3 Configuring ARP

This topic describes how to configure dynamic ARP, static ARP, and gratuitous ARP.

Context

Dynamic ARP, static ARP, and gratuitous ARP are applicable for different usage scenario. Therefore, configure them according to actual requirements.

Procedure

- Configuring dynamic ARP parameters

  Proper dynamic ARP parameter adjustment improves network reliability. The following lists optional steps.

  a. Run the **arp aging-mode** command to set the aging mode of the dynamic ARP entries.
  b. Run the **arp aging-time** command to set the aging time of the dynamic ARP entries.
  c. Run the **arp learning uni enable** command to enable user network ARP learning function for VLAN interface. By default, the system enables user network ARP learning function for VLAN interface.
  d. Run the **arp learning strict enable** command to enable dynamic ARP strict learning. By default, the system enables dynamic ARP strict learning.

- Configuring static ARP

  Static ARP applies to networks with a simple topology and high stability or networks on which information security is of high priority.

 Copyright © Huawei Technologies Co., Ltd.

a. Run the **arp** command to configure the static ARP mapping entries to specify a static mapping between an IP address and a MAC address.

   For boards on which service flows in Ethernet access mode are created by running the **service-port** command, run the **arp** *ip_addr mac_addr vlanid* **service-port** *service-portid* command to configure static ARP. If you configure static ARP by port, a configuration failure message is displayed.

- Configuring gratuitous ARP

  Gratuitous ARP reveals address conflict on a network so that ARP tables of devices can be quickly updated. This feature ensures the stability and reliability of network communication.

  a. Run the **arp gratuitous-arp period-send-mode** command to enable VLAN interfaces to periodically send gratuitous ARP packets. By default, VLAN interfaces are enabled to periodically send gratuitous ARP packets.
  b. Run the **arp conflict detect enable** command to enable the ARP conflict detection. By default, the system disables the ARP conflict detection.

----End

Result

Run the **display arp config** command to query ARP configurations.

Example

- Assume that:
- IP address of the LAN: 10.10.10.1
- Mapped MAC address: 00e0-fc01-0000
- Index of the service port: 0 (belongs to VLAN 10)

To configure static ARP by service flow, do as follows:

```
huawei(config)#arp 10.10.10.1 00e0-fc01-0000 10 service-port 0
```

- Assume that:
- IP address of the LAN: 10.10.20.1
- Mapped MAC address: 00e0-fc01-0022
- Port ID: 0/1/0 (belongs to VLAN 100)

To configure static ARP by port, do as follows:

```
huawei(config)#arp 10.10.20.1 00e0-fc01-0022 100 0/1/0
```

11.3.4 ARP Reference Standards and Protocols

The following table lists the references.

Document No.	Document Name
RFC 826	Ethernet Address Resolution Protocol
RFC 903	Reverse Address Resolution Protocol

 Copyright © Huawei Technologies Co., Ltd.

Document No.	Document Name
RFC 1027	Using ARP to Implement Transparent Subnet Gateways
RFC 1042	Standard for the Transmission of IP Datagrams over IEEE 802 Networks

11.4 ARP Proxy

ARP proxy is used to communication between users in different VLAN or in the same VLAN but isolated from each other by layer 3 forwarding.

11.4.1 Introduction to ARP proxy

Definition

When a host sends an ARP request to another host, the request is processed by the access device connected to the two hosts. This process is called ARP proxy.

Purpose

On the MA5800, ARP proxy is often used for interconnection between sub VLANs in a super VLAN.

11.4.2 Proxy ARP

Principles

The Address Resolution Protocol (ARP) is applicable only to devices on the same physical network. When a device on a physical network needs to send IP datagrams to another physical network, the gateway is used to query the routing table to implement communication between the two networks. However, routing table query consumes system resources and can affect other services. To resolve this problem, deploy proxy ARP on an intermediate device. The proxy ARP feature helps reduce system resource consumption caused by routing table queries and improves the efficiency of system processing.

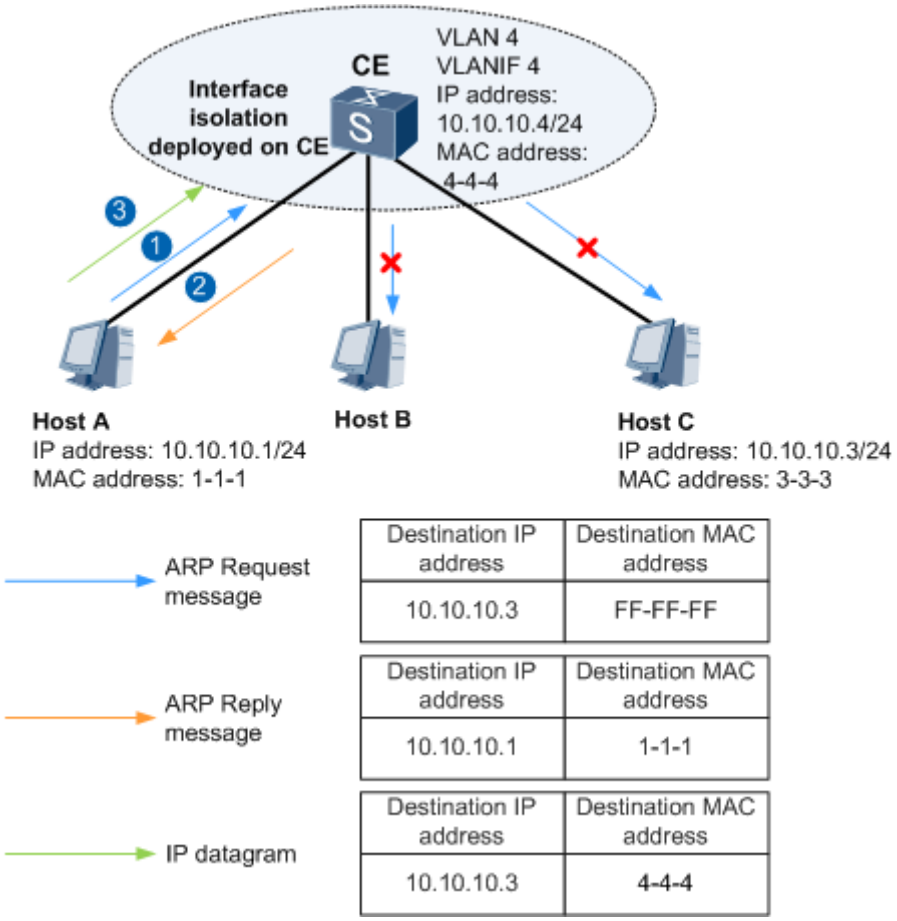
Implementation

- **Proxy ARP within a VLAN**

  Figure 11-19 illustrates how proxy ARP is implemented within a VLAN by using the communication between Host A and Host C as an example.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-19 Proxy ARP within a VLAN implementation



Host A, Host B, and Host C belong to the same VLAN. Port isolation is configured on the CE. Therefore, Host A and Host C cannot communicate at Layer 2. You can configure a VLANIF interface on the CE and enable proxy ARP within a VLAN to allow Host A and Host C to communicate.

a. Host A sends an ARP Request message for the MAC address of Host C.
b. After the CE receives the ARP Request message, the CE checks the destination IP address of the message and finds that it is not its own IP address and determines that the requested MAC address is not the MAC address of its VLANIF 4. The CE then searches its ARP table for the ARP entry indicating the mapping between the IP address and MAC address of Host C.

   - If the CE finds this ARP entry in its ARP table, the CE checks whether proxy ARP within a VLAN is enabled on it.
     - If proxy ARP within a VLAN is enabled on the CE, the CE sends the MAC address of its VLANIF 4 to Host A.
     - If proxy ARP within a VLAN is not enabled on the CE, the CE discards the ARP Request message sent by Host A.
   - If the CE does not find this ARP entry in its ARP table, the CE discards the ARP Request message sent by Host A and checks whether proxy ARP within a VLAN is enabled on it.

 Copyright © Huawei Technologies Co., Ltd.

- If proxy ARP within a VLAN is enabled on the CE, the CE broadcasts the ARP Request message with the IP address of Host C as the destination IP address within VLAN 4. After the CE receives an ARP Reply message from Host C, the CE generates an ARP entry indicating the mapping between the IP address and MAC address of Host C.
- If proxy ARP within a VLAN is not enabled on the CE, the CE does not perform any operations.

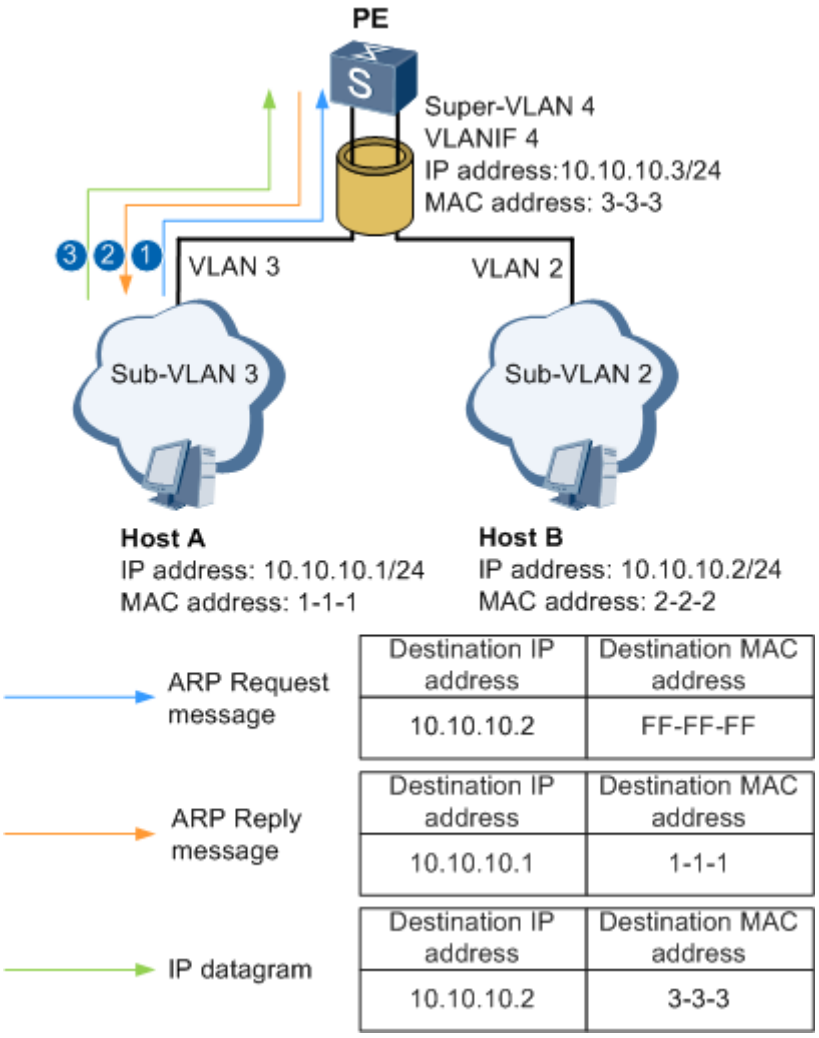
c. After Host A learns the MAC address of VLANIF 4, Host A sends IP datagrams to the CE based on this MAC address.

The CE receives the datagrams and forwards them to Host C.

- **Proxy ARP between VLANs**

  Figure 11-20 illustrates how proxy ARP is implemented between VLANs by using the communication between Host A and Host B as an example.

  Figure 11-20 Proxy ARP between VLANs implementation



Host A belongs to VLAN 3, whereas Host B belongs to VLAN 2. Therefore, Host A cannot communicate with Host B. You can configure a VLANIF

 Copyright © Huawei Technologies Co., Ltd.

interface on the PE and enable proxy ARP between VLANs on super-VLAN 4 to allow Host A and Host B to communicate.

a. Host A sends an ARP Request message for the MAC address of Host B.
b. After the PE receives the ARP Request message, the PE checks the destination IP address of the message and finds that it is not its own IP address and determines that the requested MAC address is not the MAC address of its VLANIF 4. The PE then searches its ARP table for the ARP entry indicating the mapping between the IP address and MAC address of Host B.
   - If the PE finds this ARP entry in its ARP table, the PE checks whether proxy ARP between VLANs is enabled on it.
     - If proxy ARP between VLANs is enabled on the PE, the PE sends the MAC address of its VLANIF 4 to Host A.
     - If proxy ARP between VLANs is not enabled on the PE, the PE discards the ARP Request message sent by Host A.
   - If the PE does not find this ARP entry in its ARP table, the PE discards the ARP Request message sent by Host A and checks whether proxy ARP between VLANs is enabled on it.
     - If proxy ARP within a VLAN is enabled on the CE, the CE broadcasts the ARP Request message with the IP address of Host B as the destination IP address within VLAN 2. After the PE receives an ARP Reply message from Host B, the PE generates an ARP entry indicating the mapping between the IP address and MAC address of Host B.
     - If proxy ARP between VLANs is not enabled on the PE, the PE does not perform any operations.
c. After Host A learns the MAC address of VLANIF 4, Host A sends IP datagrams to the PE based on this MAC address.

The PE receives the IP datagrams and forwards them to Host B.

Usage Scenarios

Table 11-6 Proxy ARP usage scenarios

Proxy ARP Type	Usage Scenario
Proxy ARP within a VLAN	Two hosts that need to communicate belong to the same VLAN in which user isolation is configured and the same network segment.
Proxy ARP between VLANs	Two hosts that need to communicate belong to the same network segment but different VLANs. **NOTE** In VLAN aggregation scenarios, proxy ARP between VLANs can be enabled on the VLANIF interface corresponding to the super-VLAN to implement communication between sub-VLANs.

 Copyright © Huawei Technologies Co., Ltd.

Benefits

- Proxy ARP enables a host on a network to mistakenly consider that the destination host is on the same network segment. In this manner, the details of the physical network are unavailable, and the division of the network into subnets is transparent to hosts.
- All processing related to proxy ARP is performed on a gateway, with no configuration needed on the hosts connected to it. In addition, proxy ARP affects only the ARP tables on hosts and does not affect the ARP table and routing table on a gateway.
- Proxy ARP can be used when no default gateway is configured for a host or a host cannot route messages.

11.4.3 Configuring ARP Proxy for Interworking

ARP proxy allows isolated users in a VLAN or different sub VLANs to communicate with each other. In this way, all PCs in a network segment or all voice service users connected to the AGs in the same network segment can communicate with each other. ARP proxy specially applies to triple play and enterprise users.

Context

Two users in the same network segment communicate with each other at Layer 2. To ensure user-side information security, all users connected to an ONU are isolated at Layer 2. Access services, such as the Internet access and voice services, are isolated in the same network segment. This rule also applies to FTTx networks. Specifically, the access services on the same ONU are isolated. In this situation, the users, especially enterprise users, in a LAN cannot share resources or call each other.

ARP proxy is implemented as follows: A host sends an ARP request to another one locating in the same network segment but not in the same physical network. Then, the ARP proxy-enabled device connected to the two hosts replies to this request. This allows the hosts in separate physical networks to communicate with each other.

By default, ARP proxy is disabled globally.

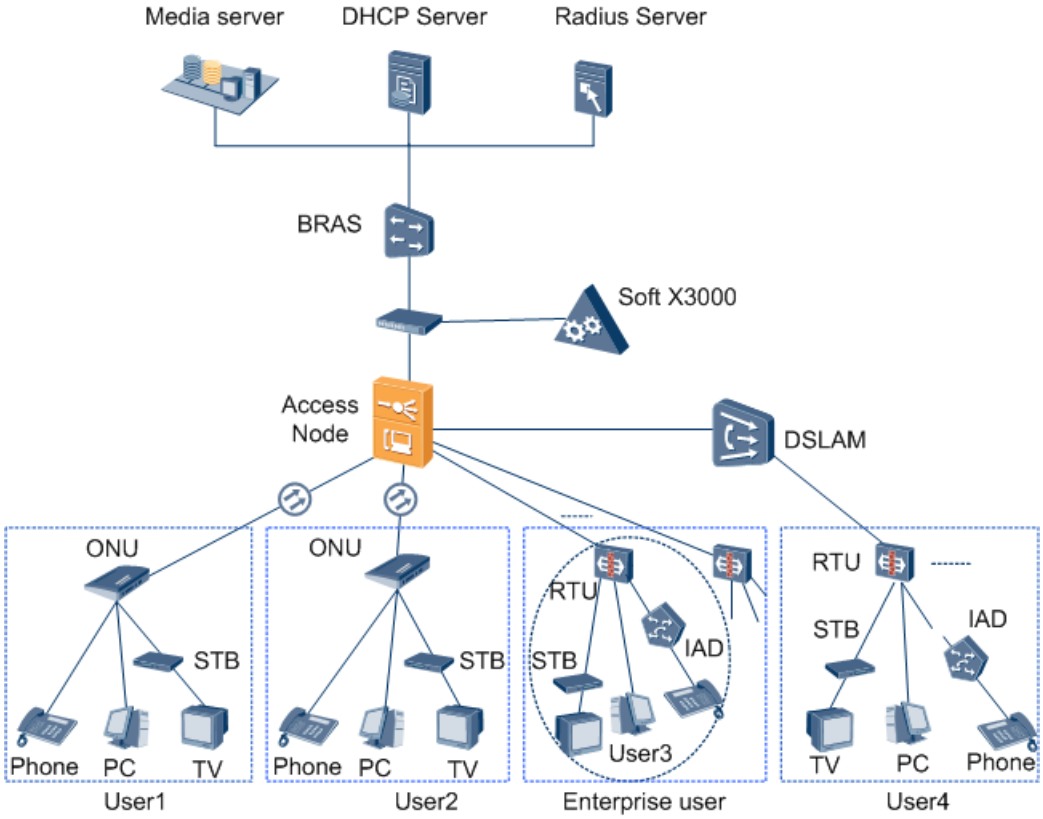
Application Context

As shown in **Figure 11-21**, ARP proxy applies to the following scenarios:

- Triple play users, users 1 and 2 in **Figure 11-21**, in the same network segment are required to communicate with each other.
- Enterprise users in the same network segment or users of different branches (enterprise branches in **Figure 11-21**) in the same enterprise are required to communicate with each other.
- Cascaded users, users 3 and 4 in **Figure 11-21**, in the same network segment are required to communicate with each other.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-21 Application scenarios of ARP proxy interworking



Prerequisite

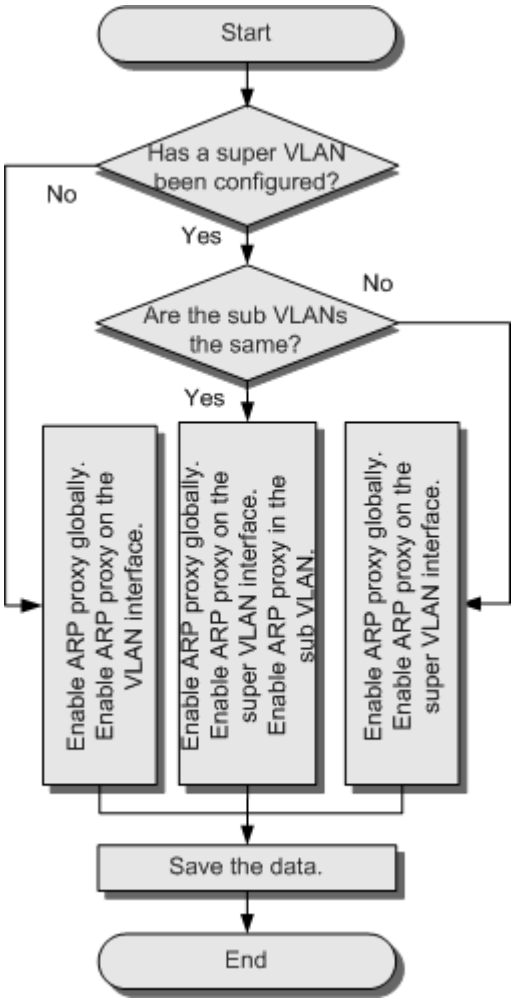
- All users to communicate with each other locate in the same VLAN or super VLAN.
- Services have been configured on the users, and the users can access the Internet and use the voice service properly.
- The IP addresses of PCs or VoIP phone sets of triple play users are in the same network segment.
- The IP addresses of PCs or IADs of enterprise users are in the same network segment.

Configuration Flow

The configuration flow varies depending on VLAN planning, as shown in **Figure 11-22**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-22 ARP proxy configuration flow



Data Plan

Table 11-7 lists the data plan for implementing the ARP proxy interworking shown in Figure 11-21.

Table 11-7 Data plan for configuring ARP proxy

Item	Data
User 1	Super VLAN: 100 Sub VLAN: smart VLAN 10
User 2	Super VLAN: 100 Sub VLAN: smart VLAN 10
Enterprise branches	Branch 1 (Smart VLAN: 50) Branch 2 (Smart VLAN: 50)
User 3	Super VLAN: 100 Sub VLAN: smart VLAN 30

 Copyright © Huawei Technologies Co., Ltd.

Item	Data
User 4	Super VLAN: 100 Sub VLAN: smart VLAN 40

Procedure

- Enable users 1 and 2 to communicate with each other.

  Both users 1 and 2 belong to super VLAN 100 and sub VLAN 10. When packet transmission in the same VLAN is isolated at Layer 2, the transmission is also isolated at Layer 3. To enable users 1 and 2 to communicate with each other, enable ARP proxy at global, super VLAN, and sub VLAN levels.

  NOTE

  - Before ARP proxy takes effect, user services must have been configured, and the users can access the Internet and use the VoIP service properly. This section does not describe service configurations. For details, see the sections for configuring the Internet access service, the VoIP service, and the triple play service.
  - The IP address of the Layer 3 interface of the super VLAN must be in the same network segment as the IP addresses of user terminals. To set the IP address of the Layer 3 interface, run the **ip address** command in the VLAN interface mode. After user terminals request for IP addresses, the DHCP server allocates the IP addresses in the same network segment to them.

  a. Configure the super VLAN and sub VLAN.

  ```
  huawei(config)#vlan 100 super
  huawei(config)#vlan 10 smart
  huawei(config)#supervlan 100 subvlan 10
  ```

  b. Enable ARP proxy globally.

  ```
  huawei(config)#arp proxy enable
  ```

  c. Enable ARP proxy on the super VLAN interface.

  ```
  huawei(config)#interface vlanif 100
  huawei(config-if-vlanif100)#arp proxy enable
  ```

  d. Enable ARP proxy in the subVLAN.

  ```
  huawei(config-if-vlanif100)#arp proxy enable subvlan 10
  huawei(config-if-vlanif100)#quit
  ```

  e. Save the data.

  ```
  huawei(config)#save
  ```

- Enable branch users of an enterprise to communicate with each other.

  No super VLAN is configured for enterprise branches, but they belong to smart VLAN 30. Therefore, the packet transmission in the same VLAN for branch users is isolated at Layer 2. To enable branch users of an enterprise to communicate with each other, enable ARP proxy at global and VLAN levels.

  a. Configure the VLAN.

  ```
  huawei(config)#vlan 30 smart
  ```

  b. Enable ARP proxy globally.

  ```
  huawei(config)#arp proxy enable
  ```

  c. Enable ARP proxy on the VLAN interface.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#interface vlanif 30
huawei(config-if-vlanif30)#arp proxy enable
huawei(config-if-vlanif30)#quit
```

d. Save the data.

```
huawei(config)#save
```

- Enable cascaded users 3 and 4 to communicate with each other.

  Cascaded users 3 and 4 belong to super VLAN 100 but different sub VLANs, sub VLANs 30 and 40, respectively. To enable cascaded users 3 and 4 to communicate with each other, enable ARP proxy globally and on the super VLAN interface.

  a. Configure the super VLAN and sub VLAN for user 3 on the access node and add cascaded port 0/3/0 to sub VLAN 40.

  NOTE

  After a cascaded port is added to a sub VLAN, this port is automatically added to the super VLAN.

  ```
  huawei(config)#vlan 100 super
  huawei(config)#vlan 30,40 smart
  huawei(config)#supervlan 100 subvlan 30
  huawei(config)#supervlan 100 subvlan 40
  huawei(config)#port vlan 40 0/3 0
  ```

  b. Enable ARP proxy globally and on the super VLAN interface on the access node.

  ```
  huawei(config)#arp proxy enable
  huawei(config)#interface vlanif 100
  huawei(config-if-vlanif100)#arp proxy enable
  ```

  c. Configure the super VLAN and sub VLAN for user 4 on the cascaded device. Then, add upstream port 0/19/0 to the super VLAN and sub VLAN, respectively.

  ```
  Cascade_dev(config)#vlan 100 super
  Cascade_dev(config)#vlan 40 smart
  Cascade_dev(config)#supervlan 100 subvlan 40
  Cascade_dev(config)#port vlan 100,40 0/19 0
  ```

----End

Result

Enabled ARP proxy supports the following functions based on service requirements:

- Triple play users, users 1 and 2 in **Figure 11-21**, in the same network segment can communicate with each other.
- Enterprise users in the same network segment or users of different branches (enterprise branches in **Figure 11-21**) in the same enterprise can communicate with each other.
- Cascaded users, users 3 and 4 in **Figure 11-21**, in the same network segment can communicate with each other.

11.4.4 ARP Proxy Reference Standards and Protocols

The following lists the reference documents of ARP proxy:

- IETF RFC1027: Using ARP to Implement Transparent Subnet Gateways

 Copyright © Huawei Technologies Co., Ltd.

11.5 DHCP Server

11.5.1 Feature Change History

Product Version	Change Description
V100R022C00	First version to support this feature.

11.5.2 Introduction to the DHCP Server

Why Is the DHCP Server Required?

A DHCP server is used to assign IP addresses in the following scenarios:

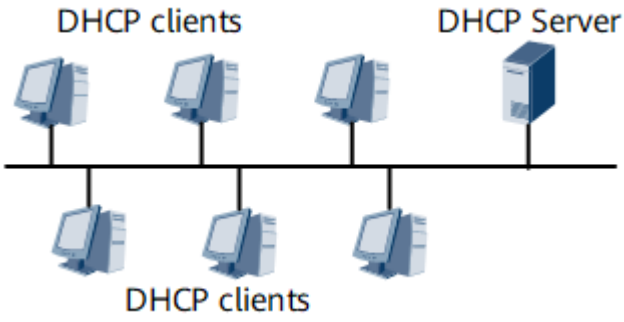
- On a large network, manual configurations take more time and efforts and bring difficulties to the centralized management over the entire network.
- The number of hosts on a network is greater than that of the available IP addresses. Thus, not every host can be assigned a fixed IP address. For example, if Internet service providers (ISPs) limit the number of concurrent network access users, many hosts must dynamically obtain IP addresses.
- Only a few hosts on the network require fixed IP addresses.

What Is a DHCP Server?

A DHCP server assigns IP addresses to clients. A client sends a message to a server to request configuration parameters, such as the IP address, subnet mask, and default gateway address. The server responds with a message carrying the requested configuration parameters. Both the request and reply messages are encapsulated in UDP packets.

Networking Scenario

As shown in Figure, a typical DHCP network consists of a DHCP server and multiple DHCP clients, such as PCs and portable computers. DHCP uses the client/server model. DHCP clients apply for configuration information from the DHCP server. The DHCP server provides configuration information including IP addresses to the DHCP clients.



 Copyright © Huawei Technologies Co., Ltd.

11.5.3 Basic Principles

A DHCP server assigns IP addresses to clients. A DHCP client sends a message to a DHCP server to request configuration parameters, such as the IP address, subnet mask, and default gateway address. The DHCP server responds with a message carrying the requested configuration parameters. Both the request and reply messages are encapsulated in UDP packets.

Modes for Interaction Between the DHCP Client and Server

To obtain a valid dynamic IP address, a DHCP client exchanges different information with a server at different stages. Generally, the DHCP client and server interact in the following modes (defined in standard protocols):

- A DHCP client accesses a network for the first time.

  When a DHCP client accesses a network for the first time, the DHCP client goes through the following stages to set up a connection with a DHCP server:

  - Discovery stage: The DHCP client searches for a DHCP server. The DHCP client broadcasts a DHCPDISCOVER message and only DHCP servers respond to the message.
  - Offer stage: Each DHCP server offers an IP address to the DHCP client. After receiving the DHCPDISCOVER message from the DHCP client, each DHCP server selects an unassigned IP address from the IP address pool, and sends a DHCPOFFER message with the leased IP address and other configurations to the DHCP client.
  - Request stage: The DHCP client selects an IP address. If multiple DHCP servers send DHCPOFFER messages to the DHCP client, the DHCP client accepts the first DHCPOFFER message it receives, and broadcasts to each DHCP server a DHCPREQUEST message carrying information about the selected IP address.
  - Acknowledgment stage: The DHCP server acknowledges the IP address that is offered. After receiving the DHCPREQUEST message, the DHCP server sends a DHCPACK message to the client. The DHCPACK message contains the offered IP address and other settings. The DHCP client then binds its TCP/IP protocol suite to the network interface card.

  Except the IP address offered by the DHCP server selected by the DHCP client, the unassigned IP addresses offered by other DHCP servers are available for other clients.

- A DHCP client extends the IP address lease.

  An IP address dynamically assigned to a DHCP client usually has a validity period. The DHCP server withdraws the IP address after the validity period expires. To continue using the IP address, the DHCP client must renew the IP address lease.

  In actual application, a DHCP client automatically sends a DHCPREQUEST message to the DHCP server to renew the IP address lease when the DHCP client is started or half the duration of the lease is remaining. If the IP address is valid, the DHCP server replies with a DHCPACK message to inform the DHCP client of the new lease.

 Copyright © Huawei Technologies Co., Ltd.

IP Address Assignment Sequence

A DHCP server assigns IP addresses to a client in the following sequence:

- IP address that is in the database of the DHCP server and is statically bound to the MAC address of the client
- IP address that is first found when the DHCP server searches the DHCP address pool for available IP addresses
- If the DHCP address pool has no available IP address, the DHCP server searches the expired IP addresses and conflicting IP addresses, and then assigns a valid IP address to the client.

Method of Preventing Repeated IP Address Assignment

To avoid address conflicts, the DHCP server pings the IP address before assigning it to a client.

The ping command checks whether a response to the ping packet is received within the specified period. If no response to the ping packet is received, the DHCP server continues to send ping packets to the IP address until the number of sent ping packets reaches the maximum limit. If there is still no response, this IP address is not in use, and the DHCP server assigns the IP address to a client. (This method is implemented based on standard protocols.)

IP Address Reservation

DHCP supports IP address reservation for clients. The reserved IP addresses must belong to the address pool. If an address in the address pool is reserved, it is no longer assignable. Addresses are usually reserved for specific clients, such as DNS and WWW servers.

11.5.4 Configuring the DHCP Server

Context

A DHCP server that uses an address pool is configured to dynamically assign IP addresses to computers that are indirectly connected to the DHCP server.

To configure the DHCP server, you need to configure the IP address of the interface, address segment, and address lease time.

Procedure

1. In the global config mode, run the **interface vlanif** command to enter the VLANIF mode.
2. In the VLANIF mode, run the **ip address** command to configure the IP address and subnet mask of the VLANIF interface.
3. In the VLANIF mode, run the **dhcp select interface** command to enable the DHCP server function.
4. In VLANIF mode, run the **dhcp server ip-range** command to configure the address pool range of the DHCP server.
5. In VLANIF mode, run the **dhcp server mask** command to configure the subnet mask of the IP address assigned by the DHCP server to a DHCP client.

 Copyright © Huawei Technologies Co., Ltd.

If the **dhcp server mask** command is used to configure a client mask, the maximum address pool range is determined by the configured mask length. Otherwise, the larger value between the 18-bit mask length and the mask length of the Layer 3 interface is used as the maximum address pool range.

6. (Optional) To prevent conflicts between the fixed IP address of the device and IP addresses in the address pool on the DHCP server, run the **dhcp server excluded-ip-address** command to configure the range of IP addresses that cannot be automatically assigned to clients from the address pool on the DHCP server. Ensure that the IP addresses that cannot be automatically assigned are not statically bound to the MAC addresses of DHCP clients using the **dhcp server static-bind** command.
7. In the VLANIF mode, run the **dhcp server lease** command to configure the lease time of the IP addresses in the address pool on the DHCP server.
8. In global config mode, run the **security anti-ipattack vlan** *vlanid* **disable** command to disable the VLAN-based anti-IP attack function.

   NOTE

   For V100R021C10, run the **security anti-ipattack disable** command to disable anti-IP attack. For V100R022C00 and later versions, run the **security anti-ipattack vlan** *vlanid* **disable** command to disable VLAN-based anti-IP attack.
9. (Optional) In VLANIF interface mode, run the **dhcp server dns-list** command to configure the IP address of the domain name system (DNS) server in the address pool of the DHCP server.
10. (Optional) In VLANIF interface mode, run the **dhcp server gateway-list** command to configure the gateway IP address assigned by the DHCP server to a DHCP client.
11. (Optional) In VLANIF interface mode, run the **dhcp server option** command to configure the user-defined options of the address pool of the DHCP server.
12. (Optional) To use a fixed IP address to ensure device stability, run the **dhcp server static-bind** command to statically bind an IP address in the address pool of the DHCP server to the MAC address of a DHCP client.

    NOTE

    If a DHCP user is online and does not age or a static user is bound by running the **dhcp server static-bind** command, you can disable the DHCP server function only after the user ages and the bound static user is deleted by running the **undo dhcp server static-bind** command. You can run the **display dhcp server interface vlanif** command to query the configurations of the DHCP server on a specified VLANIF interface.

Example

To enable the DHCP server function on VLANIF interface 2, set the address pool range of the DHCP server to 10.10.100.10 to 10.10.100.20, set the subnet mask of the IP addresses to 24, and set the IP address lease time to 2 minutes, do as follows:

```
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.10.100.1 255.255.255.0
huawei(config-if-vlanif2)#dhcp select interface
huawei(config-if-vlanif2)#dhcp server ip-range 10.10.100.10 10.10.100.20
huawei(config-if-vlanif2)#dhcp server mask 24
huawei(config-if-vlanif2)#dhcp server lease day 0 hour 0 minute 2
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-if-vlanif2)quit
huawei(config)security anti-ipattack vlan 2 disable
```

11.6 DHCPv6 Server

11.6.1 DHCPv6 Server Feature Changes

Product Version	Change Description
V100R022C10	This feature is supported for the first time.

11.6.2 Introduction to DHCPv6

Why Is a DHCPv6 Server Required?

A DHCPv6 server is used to assign IPv6 addresses in the following scenarios:

- On a large network, manual configurations take more time and efforts and bring difficulties to centralized management of the entire network.
- The number of hosts on a network is greater than that of supported IPv6 addresses. Thus, not every host can be assigned a fixed IPv6 address. For example, if Internet service providers (ISPs) limit the number of concurrent network access users, many hosts must dynamically obtain IPv6 addresses.
- Only a few hosts on the network require fixed IPv6 addresses.

What Is a DHCPv6 Server?

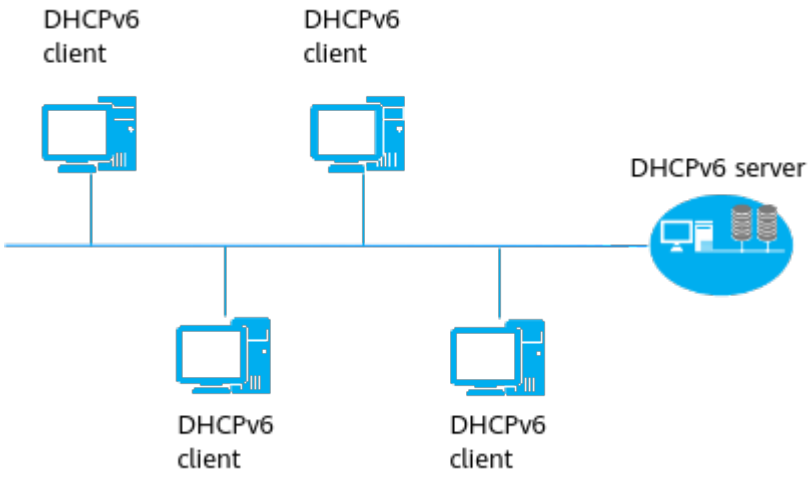
A DHCPv6 server processes address assignment, lease renewal, and address release requests from clients or relay agents, and assigns IPv6 addresses/prefixes and other network configuration parameters to clients.

Networking Scenario

As shown in **Figure 11-23**, the basic DHCPv6 architecture generally contains a DHCPv6 server and multiple DHCPv6 clients (such as PCs and laptops). DHCPv6 uses the client/server model. Clients apply for configuration information from the server. The server provides configuration information including IPv6 addresses to the clients.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-23 DHCPv6 server networking



Basic Concepts

1. Multicast address

   In DHCP, a client sends broadcast packets to locate the server. To prevent broadcast storms, IPv6 uses multicast packets instead of broadcast packets. DHCPv6 uses the following two multicast addresses:

   - FF02::1:2: multicast address of all DHCP relay agents and servers. The address is a link-local multicast address and is used for communication between a DHCPv6 client and its neighboring servers or between a DHCPv6 client and DHCPv6 relay agents. All DHCPv6 servers and relay agents are members of the multicast group corresponding to this multicast address.
   - FF05::1:3: multicast address of all DHCP servers. The address is a site-local address and is used for communication between DHCPv6 relay agents and servers within a site. All DHCPv6 servers within a site are members of this multicast group.

2. UDP port number: DHCPv6 packets are carried over UDPv6.
3. DHCP unique identifier
   - A DHCP unique identifier (DUID) identifies a DHCP device. Each server or client has a unique DUID. Servers use DUIDs to identify clients and clients use DUIDs to identify servers.
   - The DUIDs of a client and server are carried in the Client Identifier and Server Identifier options in DHCPv6 packets, respectively. The two options have the same format and are distinguished by the option-code field value. When the value of the option-code field is 1, the option is the Client Identifier option. When the value of the option-code field is 2, the option is the Server Identifier option.

11.6.3 DHCPv6 Server Principles

A DHCPv6 server assigns addresses and prefixes to clients in the following two modes:

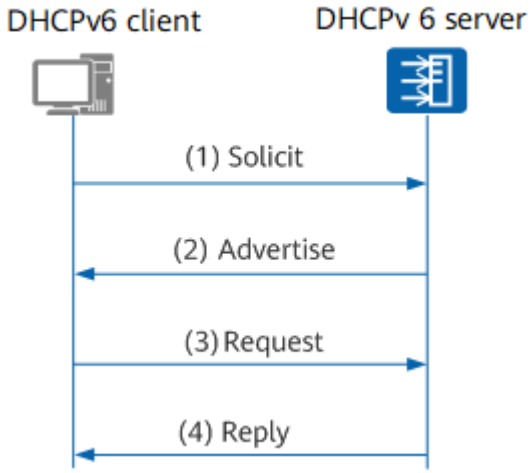
- DHCPv6 assignment with four-step interactions applies to the scenario when multiple DHCPv6 servers are deployed on a network.

 Copyright © Huawei Technologies Co., Ltd.

- DHCPv6 rapid assignment with two-step interactions applies to the scenario when only one DHCPv6 server is deployed on a network.

The process of DHCPv6 address assignment using four-step interactions is as follows:

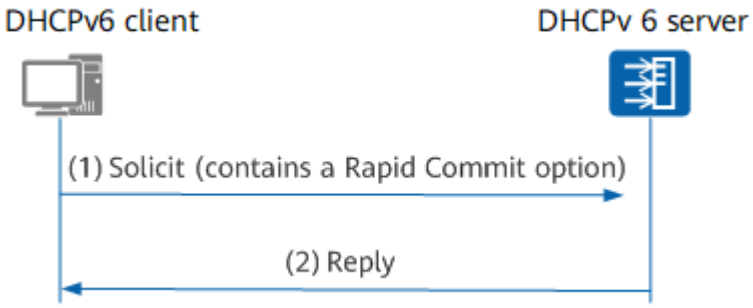
1. A DHCPv6 client sends a Solicit packet to request a DHCPv6 server to assign an IPv6 address and network configuration parameters.
2. If the Solicit packet does not carry the Rapid Commit option, or the Solicit packet carries the Rapid Commit option but the DHCPv6 server does not support the rapid assignment mode, the DHCPv6 server replies with the Advertise packet to notify the client of the address, prefix, and other network configuration parameters that can be assigned to the client.
3. If the DHCPv6 client receives Advertise packets from multiple DHCPv6 servers, the DHCPv6 client selects the DHCPv6 server with the highest priority based on the Preference option in the Advertise packets and sends a Request multicast packet to all DHCPv6 servers. The Request multicast packet carries the DUID of the selected DHCPv6 server.
4. The DHCPv6 server responds with a Reply packet, which contains the assigned address and network configuration parameters.



The process of DHCPv6 address assignment using two-step interactions is as follows:

1. A DHCPv6 client sends a Solicit packet that carries the Rapid Commit option, indicating that the client expects the server to rapidly assign an address and network configuration parameters.
2. After receiving the Solicit packet, the DHCPv6 server performs the following operations:
   - If the DHCPv6 server supports two-step rapid address assignment, it directly returns a Reply packet to assign an IPv6 address and other network configuration parameters to the client. The Reply packet also carries the Rapid Commit option.
   - If the DHCPv6 server does not support rapid address assignment, it assigns an IPv6 address/prefix and other network configuration parameters to the client through four-step interactions.

 Copyright © Huawei Technologies Co., Ltd.



11.6.4 Configuring the DHCPv6 Server

Context

If a computer on the network cannot be directly connected to a server that assigns IPv6 addresses, the computer needs to pass through other network devices. In this case, to enable the computer to dynamically obtain an IPv6 address, you need to configure a DHCPv6 server with an address pool.

To configure the DHCPv6 server, you need to configure the IPv6 address of the interface, address segment, and address lease time.

Procedure

1. In global config mode, run the **interface vlanif** command to enter VLANIF mode.
2. In VLANIF mode, run the **ipv6 enable** command to enable the IPv6 function on the interface.
3. In VLANIF mode, run the **ipv6 address** command to configure the IPv6 address of the VLAN interface.
4. In VLANIF mode, run the **dhcpv6 select interface** command to enable the DHCPv6 server function.
5. In VLANIF mode, run the **dhcpv6 server** [**allow-hint** | **preference** *preference-value* | **rapid-commit** | **unicast**] command to configure the DHCPv6 server attributes.
6. In VLANIF mode, run the **dhcpv6 server address prefix** *ipv6-prefix/ipv6-prefix-length* [**life-time** { *valid-lifetime* | **infinite** } { *preferred-lifetime* | **infinite** }] command to configure the network prefix and lifetime.
7. In VLANIF mode, run the **dhcpv6 server renew-time-percent rebind-time-percent** command to configure the renewal time and rebinding time of the IPv6 address pool.
8. In VLANIF mode, run the **dhcpv6 server excluded-address** *start-ipv6-address* [**to** *end-ipv6-address*] command to configure the range of the IPv6 addresses that cannot be automatically assigned to clients from the IPv6 address pool.

Example

In VLANIF interface 2, to set the IPv6 address 2001:db8:1::1/64 for the interface, enable DHCPv6 server address assignment, set the network prefix of the DHCPv6

 Copyright © Huawei Technologies Co., Ltd.

server to 2001:db8::/64, set the lifetime to infinite, set the percentages of the lease renewal time and rebinding time of the DHCPv6 server to the preferred lifetime of the DHCPv6 server's network prefix to 60% and 80% respectively, and exclude IPv6 addresses in the range of 2001:db8::1 to 2001:db8::10 from automatic assignment from the DHCPv6 server address pool, do as follows:

```
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ipv6 enable
huawei(config-if-vlanif2)#ipv6 address 2001:db8:1::1 64
huawei(config-if-vlanif2)#dhcpv6 select interface
huawei(config-if-vlanif2)#dhcpv6 server allow-hint
huawei(config-if-vlanif2)#dhcpv6 server address prefix 2001:db8::/64 life-time infinite infinite
huawei(config-if-vlanif2)#dhcpv6 server renew-time-percent 60 rebind-time-percent 80
huawei(config-if-vlanif2)#dhcpv6 server excluded-address 2001:db8::1 to 2001:db8::10
```

11.7 DHCP Relay

This section describes the implementation and configuration of Dynamic Host Configuration Protocol (DHCP) relay.

11.7.1 DHCP Relay Overview

What is DHCP?

Dynamic Host Configuration Protocol (DHCP) assigns IP addresses in a dynamic and centralized manner.

DHCP works in client-server (C/S) mode, including the following two processes:

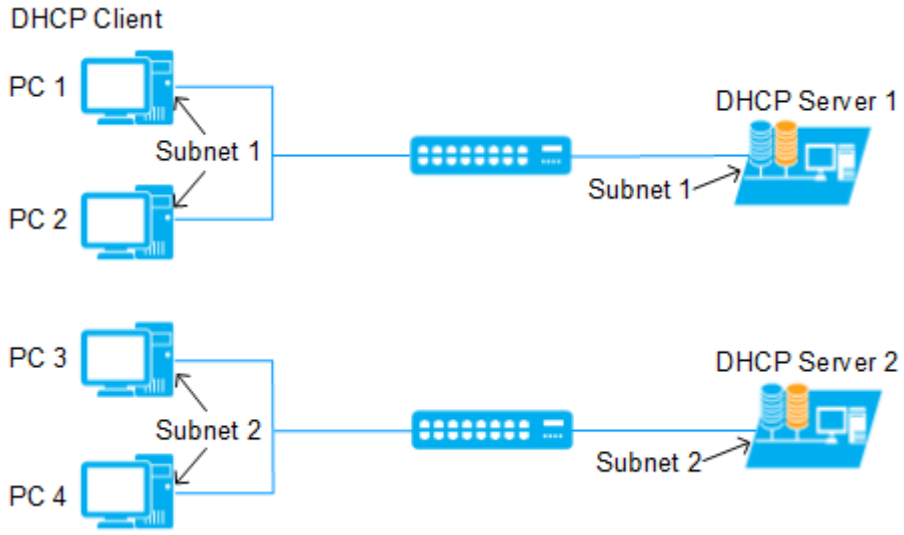
- A DHCP client sends a Request message to the DHCP server to request configuration parameters.
- The DHCP server sends a Reply message that contains information such as IP address.

Why Is DHCP Relay Needed?

As shown in **Figure 11-24**, without DHCP relay, DHCP works only when the DHCP clients and server are in the same subnet. When DHCP clients locate in different subnets, you need to deploy a DHCP server in each subnet. This is an inefficient way of using DHCP servers.

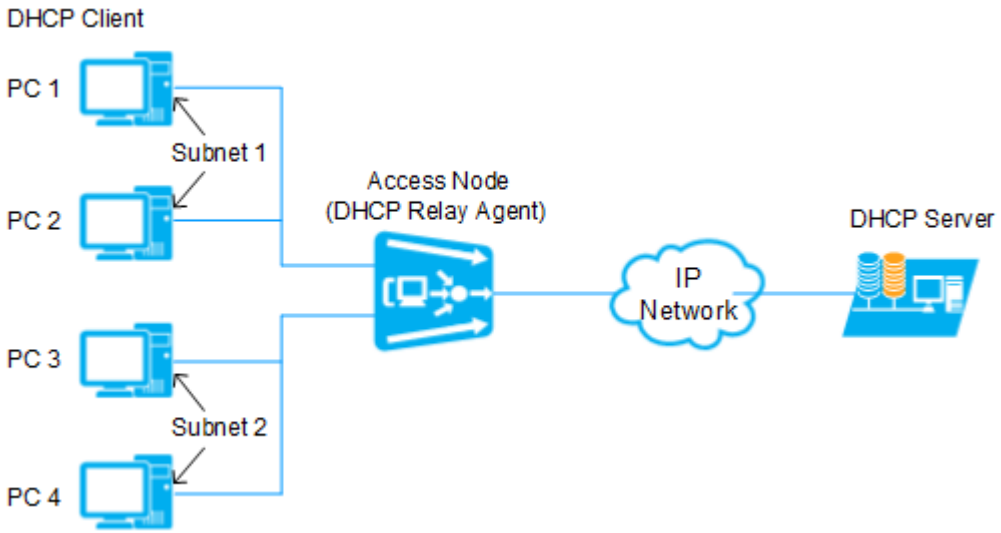
 Copyright © Huawei Technologies Co., Ltd.

Figure 11-24 Without DHCP relay



As shown in **Figure 11-25**, DHCP relay solves this problem. DHCP relay can forward DHCP packets between the DHCP server and DHCP clients across subnets. This reduces hardware costs and achieves centralized client management.

Figure 11-25 DHCP relay



- A DHCP client sends a Request message to the DHCP server to request configuration parameters.
- The DHCP server sends the client a Reply message that contains information such as IP address.

DHCP Relay Networking

As shown in **Figure 11-25**, a DHCP relay network consists of the following roles:

- DHCP client, which exchanges packets with the DHCP server to obtain an IP address.

 Copyright © Huawei Technologies Co., Ltd.

- DHCP relay agent, which forwards packets between DHCP clients and the DHCP server if they are in different subnets.
- DHCP server: assigns IP address and other network parameters to DHCP clients.

DHCP Relay Versions

DHCP relay has two versions, DHCPv4 relay and DHCPv6 relay, which apply to IPv4 and IPv6 networks, respectively.

The content in this section applies to both DHCPv4 relay and DHCPv6 relay.

11.7.2 DHCPv4 Relay Principles

DHCPv4 Layer 2 Relay Principle

If the access device between a DHCP client and the DHCP server is only a bridge device on which the routing function is unavailable or disabled, the DHCPv4 Layer 2 relay function can transparently transmit DHCP packets.

DHCPv4 Layer 3 Relay Principle

The DHCPv4 Layer 3 relay function can forward DHCP packets across subnets so that DHCP clients can obtain dynamic IP addresses from the DHCP server on another subnet. Figure 1 shows the process of the DHCPv4 L3 relay.

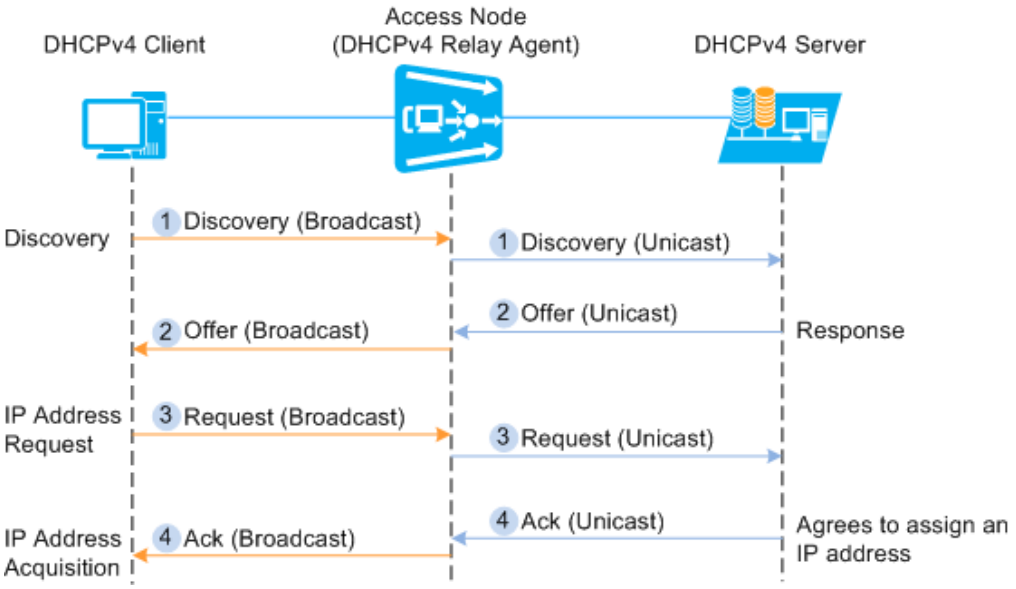
1. When DHCPv4 Relay is enabled on a DHCPv4 client, the client broadcasts a discovery message on the local network to search for the DHCPv4 server.
   - If a DHCPv4 server is available on the local network, the client can obtain an IP address from the server.
   - If no DHCPv4 server exists on the local network, the access device connected to the local network receives the Discovery message and forwards it to the DHCPv4 server on another network.
2. If the DHCPv4 server determines that the DHCPv4 client can apply for an IP address, the server returns a unicast Offer message to the access device. Upon receiving the offer message, the access device broadcasts the Offer message to the DHCPv4 clients.
3. The DHCPv4 client broadcasts a Request message to request an IP address. Upon receiving the Request message, the access device sends the Offer message to the DHCPv4 server.
4. The DHCPv4 server agrees to assign an IP address and sends an ACK message to the access device. Upon receiving the ACK message, the access device broadcasts the ACK message to configure the dynamic IP address of the DHCPv4 client.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

After DHCPv4 relay is enabled, the access device modifies the XID (a field in DHCP packet header, which is equivalent to the sequence number of a DHCP packet) of DHCP packets sent by users. As a result, the XIDs in the DHCP packets sent by DHCPv4 clients are different from those in the DHCP packet received by the DHCPv4 server. Generally, the DHCPv4 server does not verify the XID. Therefore, services are not affected. If an operator adds information to the XID field of the packets sent by DHCPv4 clients for verification on the DHCPv4 server, the verification may fail because the added information is not defined in the protocol. In this case, services may be interrupted.

Figure 11-26 DHCPv4 Layer 3 relay working process



Concerning the DHCPv4 Layer 3 relay principle, you need to understand the following two questions.

1. How does a DHCPv4 relay agent convert broadcast packets into unicast packets?

   The DHCPv4 relay agent knows the IP address of the DHCPv4 server. Therefore, it changes the broadcast IP addresses in Discovery and Request messages received from the clients to the destination IP address of the DHCPv4 server.

2. How does a DHCPv4 server know which address pool is used to allocate IP addresses to clients?

   Upon receiving the Discovery messages from the clients, the DHCPv4 relay agent adds the gateway address, for example, 10.1.1.1, to the messages. After receiving the Discovery messages, the DHCPv4 server determines that the requests are sent from network 10.1.1.0/24, and then assigns IP addresses from the range of 10.1.1.2–10.1.1.254.

DHCPv4 Relay and DHCP Option 82

The DHCPv4 relay function has no authentication or security mechanisms. In actual applications, there are many security problems, such as excessive broadcast packets and IP exhaustion attacks. The DHCPv4 relay agent inserts Option 82 into

 Copyright © Huawei Technologies Co., Ltd.

DHCPv4 packets. The DHCPv4 server can determine the validity of the DHCPv4 clients based on the Option 82 information, ensuring user access security. When the DHCP Option 82 function is enabled, the DHCPv4 relay agent adds or removes Option 82 in the received DHCPv4 packets regardless of DHCPv4 L2 Relay or DHCPv4 L3 Relay. For details, see **DHCP Option 82**.

11.7.3 DHCPv6 Relay Principle

Principle of DHCPv6 L2 Relay

By default, the system checks the IPv6 addresses of all DHCPv6 packets according to the RFC6221 standard. If an IPv6 address meets the requirement, the corresponding packet is relayed. If an IPv6 address does not meet the requirement, the corresponding packet is directly forwarded and is not relayed. The check criteria are as follows:

- DHCPv6 request packet: The destination address is All_DHCP_Relay_Agents_and_Servers (ff02::1:2).
- DHCPv6 response packet: The source and destination addresses are both the link-local address (LLA).

The DHCPv4 Layer 2 relay working process is as follows:

1. When a DHCPv6 relay agent receives a message that needs to be processed using L2 relay, the relay agent constructs a new Relay-forward message, regardless of the original message type. The relay agent copies the IP address of the original message to the peer-address field of the new Relay-forward message, and copies the entire original DHCP message (excluding the IP header and UDP header) to the Relay Message Option (Option 9) of the new Relay-forward message.
2. If the original message originates from a DHCPv6 client, the DHCPv6 relay agent initializes the hop-count field of the Relay-forward message to 0. If the original message is a Relay-forward message that originates from another DHCPv6 relay agent, the relay agent first checks the hop-count field of the message. If the value of the field is greater than or equal to 32, the relay agent discards the message; if the value is smaller than 32, the relay agent adds 1 to the value and uses the new value as the value of the hop-count field of the new Relay-forward message.
3. The DHCPv6 relay agent invariably sets the link-address Field of the Relay-forward message to an unspecified address (::), and invariably includes the interface-id field (Option 18) and remote-id field (Option 37) in the Relay-forward message.
4. In the downstream direction, the Relay Message Option (Option 9) of the original Relay-reply message is extracted, and the content of the Relay Message Option (Option 9) is forwarded as a new downstream message to the DHCPv6 client.

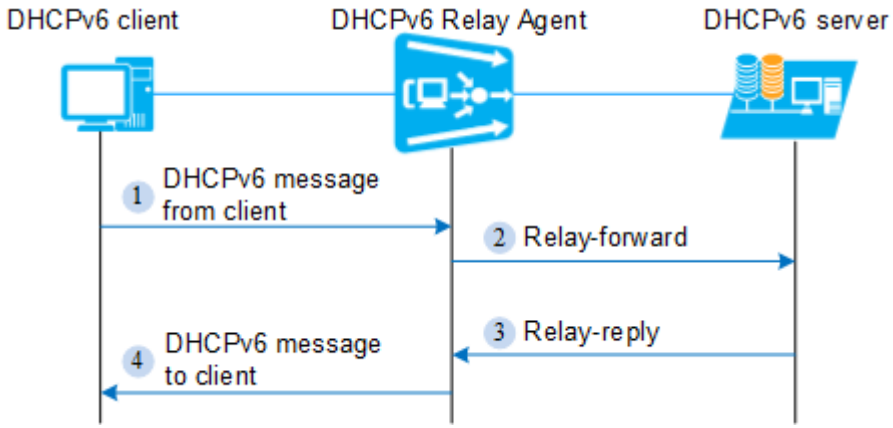
Principle of DHCPv6 L3 Relay

The processing flows between the DHCPv6 client and server are the same when the IPv6 address, IPv6 prefix and other network configuration parameters are dynamically obtained using or not using the DHCPv6 relay agent. **Figure 11-27** illustrates only the forwarding process when DHCPv6 relay is used.

 Copyright © Huawei Technologies Co., Ltd.

The MA5800 supports DHCPv6 option 53. DHCPv6 option 53 is Relay-ID option which carries the DUID information of DHCPv6 L3 relay agent.

Figure 11-27 Working process with DHCPv6 relay



1. The DHCPv6 client sends a request message to the multicast address FF02::1:2 of all DHCPv6 servers and relay agents.
2. After receiving the request message, the DHCPv6 relay agent encapsulates the message into the Relay Message Option of the Relay-forward message and sends the Relay-forward message to the DHCPv6 server.
3. The DHCPv6 server parses the Relay-forward message and obtains the client's request message. Then, the server selects an IPv6 address and other parameters for the client, constructs a reply message, encapsulates the reply message into the Relay Message Option of the Relay-reply message, and sends the Relay-reply message to the DHCPv6 relay agent.
4. The DHCPv6 relay agent parses the Relay-reply message, obtains the server's reply message, and then forwards the message to the DHCPv6 client.
5. The DHCPv6 client implements network configuration according to the IPv6 address, IPv6 prefix, and other parameters contained in the reply message.

NOTE

A DHCPv6 server allocates both IPv6 addresses and IPv6 prefixes to DHCPv6 clients. After obtaining an IPv6 prefix allocated by the DHCPv6 server, a DHCPv6 client sends a Router Advertisement (RA) message containing the IPv6 prefix to the network in which it is located. Using the prefix, hosts in the network can automatically configure their IPv6 addresses.

11.7.4 Configuring DHCPv4 Relay

The MA5800 can work in DHCPv4 Layer 2 relay or DHCPv4 Layer 3 relay mode to forward packets between DHCPv4 clients and the server across networks.

- By default, the MA5800 works in the Layer 2 DHCP relay mode. In this mode, the MA5800 transparently transmits the DHCPv4 packets from users.
- When the MA5800 works in Layer 3 mode, DHCPv4 server groups need to be configured. A DHCPv4 server group contains one or more DHCPv4 servers. A DHCPv4 server can be added to multiple DHCPv4 server groups at the same time.

 Copyright © Huawei Technologies Co., Ltd.

DHCPv4 Layer 3 relay mode can be configured on the MA5800 to select a DHCPv4 server group. The following DHCPv4 relay modes are supported:

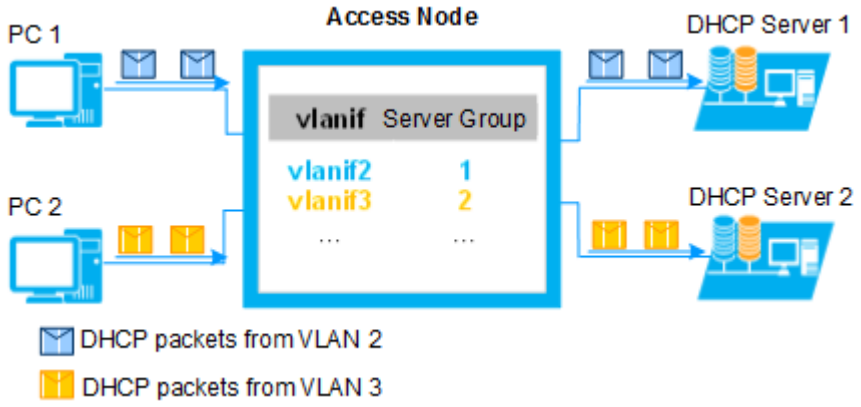
Standard Mode

Figure 11-28 shows the standard mode, which is the default mode and is applicable when you need to specify a DHCPv4 server group for users in different VLANs. In the preceding figure, the two servers belong to two server groups. In actual deployment, the server groups can be mapped to one or multiple servers. The DHCPv4 relay agent selects a DHCPv4 server group based on the addresses of the primary Layer 3 VLAN interfaces, which are the gateway addresses of DHCPv4 relay agents. This mode is the most common and simplest DHCPv4 relay mode. The disadvantage of this mode is that different services in the same VLAN cannot be distinguished.

To use this mode, you must bind Layer 3 VLAN interfaces to DHCP server groups. The configuration procedure is as follows:

Set the working mode of DHCPv4 relay to standard, configure DHCPv4 server groups, and bind VLANs to DHCPv4 server groups.

Figure 11-28 Standard mode of DHCPv4 relay



Option 60 Mode

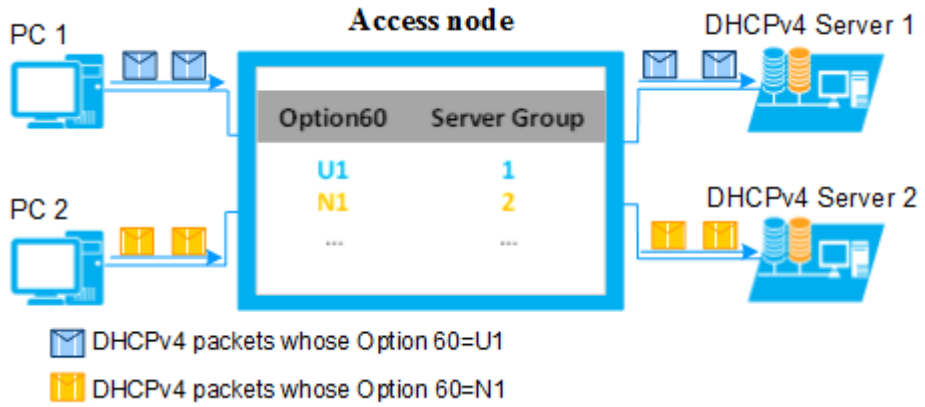
As shown in Figure 2 DHCPv4 Relay Option 60 Mode, this mode applies when you need to specify DHCPv4 server groups for users in different domains indicated by Option 60 strings. The DHCPv4 relay agent selects a DHCPv4 server group based on Option 60 strings (domain names) in the received DHCPv4 packets. This mode is also commonly used to distinguish services in the same VLAN through the Option 60 strings.

To use this mode, you must configure domain names through the Option 60 strings and bind the domain names to DHCPv4 server groups. The configuration procedure is as follows:

Set the DHCPv4 relay working mode to Option 60, configure DHCPv4 server groups, create domains using the Option 60 strings, and bind the Option 60 strings to DHCPv4 server groups.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-29 DHCPv4 relay Option 60 mode



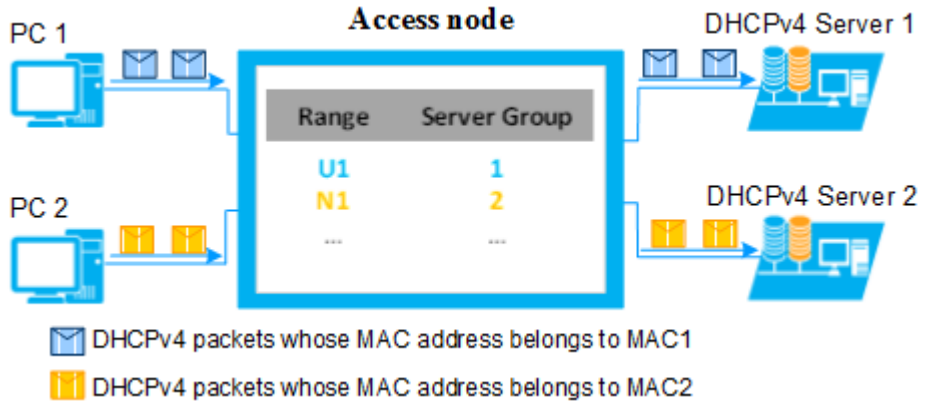
MAC Address Range Mode

As shown in **Figure 11-30**, this mode applies when you need to specify DHCPv4 server groups for users in different MAC address ranges. The DHCPv4 relay agent selects a DHCPv4 server group based on the source MAC address range of the received DHCPv4 packets. In this mode, services in the same VLAN are distinguished based on the MAC address range.

To use this mode, you must configure MAC address ranges and bind them to DHCPv4 server groups. The configuration procedure is as follows:

Set the DHCPv4 relay working mode to mac-range, configure DHCPv4 server groups, define MAC address ranges, and bind the MAC address ranges to DHCPv4 server groups.

Figure 11-30 DHCPv4 relay MAC address range mode



11.7.4.1 Configuring the Standard Mode

The standard mode is used to specify a DHCP server group for users of a VLAN that contains service ports created on the MA5800.

Prerequisites

The user VLAN and network VLAN have been configured. For details about how to configure a VLAN, see **Configuring a VLAN**.

 Copyright © Huawei Technologies Co., Ltd.

Procedure

Step 1 Configure the DHCP forwarding mode.

The DHCP forwarding mode can be configured in global config mode or VLAN service profile mode.

- In global config mode, run the **dhcp mode layer-3 standard** command to configure the DHCP Layer 3 relay mode to **standard**. If you have selected the **vlan** keyword and specified a VLAN ID, the standard DHCP Layer 3 relay mode takes effect only in the specified VLAN.
- In VLAN service profile mode, perform the following operations to configure the DHCP forwarding mode in the user VLAN:
  a. Run the **vlan service-profile** command to create a VLAN service profile.
  b. Run the **dhcp mode layer-3 standard** command to configure the DHCP Layer 3 relay mode to **standard**.
  c. Run the **commit** command to make the profile configuration take effect.
  d. Run the **quit** command to quit the VLAN service profile mode.
  e. Run the **vlan bind service-profile** command to bind the VLAN service profile created in **Step 1.a** to the VLAN.

Step 2 Configure a DHCP server group.

1. In global config mode, run the **dhcp-server** command to create a DHCP server group.

   igroup-number and *ip-addr* are keywords in the **dhcp-server** command.

   – *igroup-number*: number that uniquely identifies a DHCP server group.

     Before specifying a number for a DHCP server group, check whether the number is unique among the DHCP server group numbers that have been configured on the MA5800. To do so, run the **display dhcp-server all-group** command on the MA5800.

   – *ip-addr*: IP address of a DHCP server in a DHCP server group. A maximum of four IP addresses can be specified for this parameter.

     NOTE

     Ensure that the IP address configured in this parameter is the same as the IP address of the network-side DHCP server.

2. (Optional) Run the **dhcp server mode** command to configure the working mode of the DHCP servers in the DHCP server group.

   The DHCP servers in a DHCP server group can work in load sharing or active/standby mode. By default, the DHCP servers work in load sharing mode.

Step 3 Bind the VLAN to the DHCP server.

1. In global config mode, run the **interface vlanif** command to create a VLAN Layer 3 interface.

   Ensure that the value of the **VLANID** keyword in the **interface vlanif** command is the same as the ID of the VLAN that has been created.

2. In VLAN interface (VLANIF) mode, run the **ip address** command to configure the IP address of the VLAN Layer 3 interface.

 Copyright © Huawei Technologies Co., Ltd.

After the configuration, all IP packets in the VLAN use this IP address as the source IP address for Layer 3 forwarding.

NOTE

- If all the devices between the MA5800 and the DHCP server work at Layer 2, the IP address of the VLAN Layer 3 interface must be in the same network segment as the IP address of the DHCP server.
- If a device between the MA5800 and the DHCP server works at Layer 3, the IP address of the VLAN Layer 3 interface can be in a network segment different from that of the DHCP server. In this case, ensure that the DHCP server is reachable from the VLAN Layer 3 interface.

3. Versions earlier than V100R020C00, in VLANIF mode, run the **dhcp-server** command to bind the VLAN to the DHCP server.

   In V100R020C00 and later versions, in VLANIF-UAM mode, run the **dhcp-server** command to bind the VLAN to the DHCP server.

   NOTE

   In V100R020C00 and later versions, the DHCP relay service port is disabled by default. Run the **sysman service dhcp-relay enable** command to enable the DHCP relay service.

   This command requires the *group-number*. Ensure that the *group-number* value is the same as the number of the DHCP server group.

----End

Example

The following is an example of the configurations used to enable DHCP relay on the MA5800 so that DHCP clients in VLAN 2 can obtain IP addresses from DHCP server group 1:

- The DHCP clients in VLAN 2 are in a network segment different from DHCP server group 1.
- The IP address of the Layer 3 interface in VLAN 2 is 10.1.1.101/24.
- DHCP server group 1 contains two DHCP servers working in active/standby mode.
  - The IP address of the active server is 10.1.1.9.
  - The IP address of the standby server is 10.1.1.10.
- The maximum response time of the DHCP servers is 20s.
- The maximum number of DHCP server response timeout times is 10.

```
Versions earlier than V100R020C00
huawei(config)#dhcp mode layer-3 standard
huawei(config)#dhcp server mode backup 20 10
huawei(config)#dhcp-server 1 ip 10.1.1.9 10.1.1.10
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.1.1.101 24
huawei(config-if-vlanif2)#dhcp-server 1
In V100R020C00 and later versions
huawei(config)#dhcp mode layer-3 standard
huawei(config)#dhcp server mode backup 20 10
huawei(config)#dhcp-server 1 ip 10.1.1.9 10.1.1.10
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.1.1.101 24
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-if-vlanif2)#quit
huawei(config)sysman service dhcp-relay enable
huawei(config)#interface vlanif-uam 2
huawei(config-if-vlanif-uam-2)#dhcp-server 1
```

11.7.4.2 Configuring the DHCP Option 60 Mode

The DHCP Option 60 mode is used to specify a DHCP server group for DHCP Option 60 users.

Prerequisites

- The user VLAN and network VLAN have been configured. For details about how to configure a VLAN, see **Configuring a VLAN**.
- Before configuring the DHCP Option 60 mode, ensure that the Option 60 domain name is available.

Context

If the MA5800 provisions multiple services, such as the multicast video and VoIP services, the services may be provided by different service providers. These service providers use different DHCP servers or different relay IP addresses of the same DHCP server to allocate IP addresses to DHCP users. Therefore, the DHCP Option 60 mode needs to be configured for the users in different DHCP Option 60 domains to apply for IP addresses.

Procedure

Step 1 Configure the DHCP forwarding mode.

The DHCP forwarding mode can be configured in global config mode or VLAN service profile mode.

- In global config mode, run the **dhcp mode layer-3 option60** command to configure the DHCP Layer 3 relay mode to **option60**. If you have selected the **vlan** keyword and specified a VLAN ID, the Option 60 DHCP Layer 3 relay mode takes effect only in the specified VLAN.
- In VLAN service profile mode, perform the following operations to configure the DHCP forwarding mode in the user VLAN:
  a. Run the **vlan service-profile** command to create a VLAN service profile.
  b. Run the **dhcp mode layer-3 option60** command to configure the DHCP Layer 3 relay mode to **option60**.
  c. Run the **commit** command to make the profile configuration take effect.
  d. Run the **quit** command to quit the VLAN service profile mode.
  e. Run the **vlan bind service-profile** command to bind the VLAN service profile created in **Step 1.a** to the VLAN.

Step 2 Configure a DHCP server group.

1. In global config mode, run the **dhcp-server** command to create a DHCP server group.

   igroup-number and *ip-addr* are keywords in the **dhcp-server** command.

 Copyright © Huawei Technologies Co., Ltd.

- *igroup-number*: number that uniquely identifies a DHCP server group.

  Before specifying a number for a DHCP server group, check whether the number is unique among the DHCP server group numbers that have been configured on the MA5800. To do so, run the **display dhcp-server all-group** command on the MA5800.
- *ip-addr*: IP address of a DHCP server in a DHCP server group. A maximum of four IP addresses can be specified for this parameter.

  NOTE

  Ensure that the IP address configured in this parameter is the same as the IP address of the network-side DHCP server.

2. (Optional) Run the **dhcp server mode** command to configure the working mode of the DHCP servers in the DHCP server group.

   The DHCP servers in a DHCP server group can work in load sharing or active/standby mode. By default, the DHCP servers work in load sharing mode.

Step 3 In global config mode, run the **dhcp domain** command to create a DHCP Option 60 domain.

Configure the domain name based on the type of the terminal connected to the MA5800. For example, if the terminal connected to the MA5800 is a DHCP client running the Windows 98, Windows 2000, Windows XP, or Windows NT OS, the domain name must be **msft**.

Step 4 In DHCP Option 60 mode, run the **dhcp-server** command to bind the DHCP Option 60 domain to the DHCP server group.

After the configuration, the DHCP server group serves all DHCP clients in the DHCP Option 60 domain.

NOTE

In V100R020C00 and later versions, the DHCP relay service port is disabled by default. Run the **sysman service dhcp-relay enable** command to enable the DHCP relay service.

Step 5 Configure the gateway address of the DHCP Option 60 domain.

1. In global config mode, run the **interface vlanif** command to create a VLAN Layer 3 interface.

   Ensure that the value of the **VLANID** keyword in the **interface vlanif** command is the same as the ID of the VLAN that has been created.
2. In VLANIF mode, run the **ip address** command to configure the IP address of the VLAN Layer 3 interface.

   After the configuration, all IP packets in the VLAN use this IP address as the source IP address for Layer 3 forwarding.

   NOTE

   - If all the devices between the MA5800 and the DHCP server work at Layer 2, the IP address of the VLAN Layer 3 interface must be in the same network segment as the IP address of the DHCP server.
   - If a device between the MA5800 and the DHCP server works at Layer 3, the IP address of the VLAN Layer 3 interface can be in a network segment different from that of the DHCP server. In this case, ensure that the DHCP server is reachable from the VLAN Layer 3 interface.

 Copyright © Huawei Technologies Co., Ltd.

3. Versions earlier than V100R020C00, in VLANIF mode, run the **dhcp domain gateway** command to configure the gateway address of the DHCP Option 60 domain.

   In V100R020C00 and later versions, in VLANIF-UAM mode, run the **dhcp domain gateway** command to configure the gateway address of the DHCP Option 60 domain.

   Ensure that this gateway address is one of the IP addresses of the VLAN Layer 3 interface. Different gateways can be configured for different DHCP Option 60 domains under the same VLAN Layer 3 interface. Therefore, the DHCP Option 60 users connected to different DHCP servers can be differentiated based on their DHCP Option 60 domain name.

----End

Example

The following is an example of the configurations used to enable DHCP relay on the MA5800 so that DHCP clients in VLAN 2 can obtain IP addresses from DHCP server group 1.

- The DHCP clients run the Windows 98, Windows 2000, Windows XP, or Windows NT OS.
- The DHCP clients in VLAN 2 are in a network segment different from DHCP server group 1.
- The IP address of the Layer 3 interface in VLAN 2 is 10.1.2.1/24.
- DHCP server group 1 contains two DHCP servers working in load sharing mode.
  - The IP address of the active server is 10.10.10.10.
  - The IP address of the standby server is 10.10.10.11.

```
Versions earlier than V100R020C00
huawei(config)#dhcp mode layer-3 Option60
huawei(config)#dhcp-server 1 ip 10.10.10.10 10.10.10.11
huawei(config)#dhcp domain msft
huawei(config-dhcp-domain-msft)#dhcp-server 1
huawei(config-dhcp-domain-msft)#quit
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.1.2.1 24
huawei(config-if-vlanif2)#dhcp domain msft gateway 10.1.2.1
In V100R020C00 and later versions
huawei(config)#dhcp mode layer-3 Option60
huawei(config)#dhcp-server 1 ip 10.10.10.10 10.10.10.11
huawei(config)#dhcp domain msft
huawei(config-dhcp-domain-msft)#dhcp-server 1
huawei(config-dhcp-domain-msft)#quit
huawei(config)sysman service dhcp-relay enable
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.1.2.1 24
huawei(config-if-vlanif2)#quit
huawei(config)#interface vlanif-uam 2
huawei(config-if-vlanif-uam-2)#dhcp domain msft gateway 10.1.2.1
```

11.7.4.3 Configuring the DHCP MAC Address Segment Mode

The DHCP MAC address segment mode is used to specify a DHCP server group for the users in a MAC address segment.

 Copyright © Huawei Technologies Co., Ltd.

Prerequisites

The user VLAN and network VLAN have been configured. For details about how to configure a VLAN, see **Configuring a VLAN**.

Context

The devices on a network may be from different vendors. Each vendor has a consistent MAC address segment and all devices from a vendor use the vendor's MAC address segment. Configure the DHCP MAC address segment mode for the devices to obtain IP addresses from the DHCP server.

The MA5800 selects a DHCP server group based on a MAC address segment. After a MAC address segment is configured, all DHCP clients in the MAC address segment can obtain IP addresses from the DHCP server group.

Procedure

Step 1 Configure the DHCP forwarding mode.

The DHCP forwarding mode can be configured in global config mode or VLAN service profile mode.

- In global config mode, run the **dhcp mode layer–3 mac-range** command to configure the DHCP Layer 3 relay mode to **mac-range**. If you have selected the **vlan** keyword and specified a VLAN ID, the DHCP MAC address segment mode takes effect only in the specified VLAN.
- In VLAN service profile mode, perform the following operations to configure the DHCP forwarding mode in the user VLAN:
  a. Run the **vlan service-profile** command to create a VLAN service profile.
  b. Run the **dhcp mode layer–3 mac-range** command to configure the DHCP Layer 3 relay mode to **mac-range**.
  c. Run the **commit** command to make the profile configuration take effect.
  d. Run the **quit** command to quit the VLAN service profile mode.
  e. Run the **vlan bind service-profile** command to bind the VLAN service profile created in **Step 1.a** to the VLAN.

Step 2 Configure a DHCP server group.

1. In global config mode, run the **dhcp-server** command to create a DHCP server group.

   igroup-number and *ip-addr* are keywords in the **dhcp-server** command.

   – *igroup-number*: number that uniquely identifies a DHCP server group.

     Before specifying a number for a DHCP server group, check whether the number is unique among the DHCP server group numbers that have been configured on the MA5800. To do so, run the **display dhcp-server all-group** command on the MA5800.

   – *ip-addr*: IP address of a DHCP server in a DHCP server group. A maximum of four IP addresses can be specified for this parameter.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

Ensure that the IP address configured in this parameter is the same as the IP address of the network-side DHCP server.

2. (Optional) Run the **dhcp server mode** command to configure the working mode of the DHCP servers in the DHCP server group.

   The DHCP servers in a DHCP server group can work in load sharing or active/standby mode. By default, the DHCP servers work in load sharing mode.

Step 3 Specify a MAC address segment.

1. In global config mode, run the **dhcp mac-range** command to create a MAC address segment.

   range-name is the MAC address segment name, which is used only for commenting on the MAC address segment.

2. In MAC address segment mode, run the **mac-range mac-address-start to mac-address-end** command to specify the range of the MAC address segment.

Step 4 In MAC address segment mode, run the **dhcp-server** command to bind the MAC address segment to the DHCP server.

NOTE

In V100R020C00 and later versions, the DHCP relay service port is disabled by default. Run the **sysman service dhcp-relay enable** command to enable the DHCP relay service.

Step 5 Configure the gateway address of the MAC address segment.

1. In global config mode, run the **interface vlanif** command to create a VLAN Layer 3 interface.

   Ensure that the value of the **VLANID** keyword in the **interface vlanif** command is the same as the ID of the VLAN that has been created.

2. In VLANIF mode, run the **ip address** command to configure the IP address of the VLAN Layer 3 interface.

   After the configuration, all IP packets in the VLAN use this IP address as the source IP address for Layer 3 forwarding.

   NOTE

   - If all the devices between the MA5800 and the DHCP server work at Layer 2, the IP address of the VLAN Layer 3 interface must be in the same network segment as the IP address of the DHCP server.
   - If a device between the MA5800 and the DHCP server works at Layer 3, the IP address of the VLAN Layer 3 interface can be in a network segment different from that of the DHCP server. In this case, ensure that the DHCP server is reachable from the VLAN Layer 3 interface.

3. Versions earlier than V100R020C00, in VLANIF mode, run the **dhcp mac-range gateway** command to configure the gateway address of the MAC address segment.

   In V100R020C00 and later versions, in VLANIF-UAM mode, run the **dhcp mac-range gateway** command to configure the gateway address of the MAC address segment.

 Copyright © Huawei Technologies Co., Ltd.

Ensure that this gateway address is one of the IP addresses of the VLAN Layer 3 interface. Different gateways can be configured for different MAC address segments under the same VLAN Layer 3 interface. Therefore, the users connected to different DHCP servers can be differentiated based on their DHCP MAC address segment.

----End

Example

The following is an example of the configurations used to enable DHCP relay on the MA5800 so that users in the MAC address segment ranging from 00e0-fc00-0001 to 00e0-fc00-0100 in VLAN 2 can obtain IP addresses from DHCP server group 1.

- The DHCP clients in VLAN 2 are in a network segment different from DHCP server group 1.
- The IP address of the Layer 3 interface in VLAN 2 is 10.1.2.1/24.
- DHCP server group 1 contains two DHCP servers working in load sharing mode.
  - The IP address of the active server is 10.10.10.10.
  - The IP address of the standby server is 10.10.10.11.

```
Versions earlier than V100R020C00
huawei(config)#dhcp mode layer-3 mac-range
huawei(config)#dhcp-server 1 ip 10.10.10.10 10.10.10.11
huawei(config)#dhcp mac-range huawei
huawei(config-mac-range-huawei)#mac-range 00e0-fc00-0001 to 00e0-fc00-0100
huawei(config-mac-range-huawei)#dhcp-server 1
huawei(config-mac-range-huawei)#quit
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.1.2.1 24
huawei(config-if-vlanif2)#dhcp mac-range huawei gateway 10.1.2.1
In V100R020C00 and later versions
huawei(config)#dhcp mode layer-3 mac-range
huawei(config)#dhcp-server 1 ip 10.10.10.10 10.10.10.11
huawei(config)#dhcp mac-range huawei
huawei(config-mac-range-huawei)#mac-range 00e0-fc00-0001 to 00e0-fc00-0100
huawei(config-mac-range-huawei)#dhcp-server 1
huawei(config-mac-range-huawei)#quit
huawei(config)sysman service dhcp-relay enable
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.1.2.1 24
huawei(config-if-vlanif2)#quit
huawei(config)#interface vlanif-uam 2
huawei(config-if-vlanif-uam-2)#dhcp mac-range huawei gateway 10.1.2.1
```

11.7.5 Differences Between DHCPv4 and DHCPv6 Configurations

Dynamic Host Control Protocol version 6 (DHCPv6) is a DHCP protocol for IPv6. This topic describes the differences between DHCPv4 and DHCPv6 regarding function specifications and commands. Before configuring DHCPv6 services, it is recommended that you familiarize yourself with the procedures and principles of configuring DHCPv4 services and be aware of the differences between DHCPv4 and DHCPv6 configurations.

 Copyright © Huawei Technologies Co., Ltd.

Context

The differences between DHCPv4 and DHCPv6 configurations are as follows:

- When working in Layer 3 forwarding mode, DHCPv6 supports both the standard mode and option 16 mode. You can run the **dhcpv6 mode { layer-2 | layer-3 [option16] }** command to configure the DHCPv6 working mode. DHCPv4 supports the standard, media access control (MAC) address segment, and DHCP option 60 modes. You can run the **dhcp mode { layer-2 | layer-3 { mac-range | option60 | standard } }** command to configure the DHCPv4 working mode.
- DHCPv6 does not support DHCP proxy.
- Regarding commands, DHCPv4 uses **dhcp** while DHCPv6 uses **dhcpv6** as the command word. For example, the **dhcp-server** command is used to configure DHCPv4 server groups while the **dhcpv6-server** command is used to configure DHCPv6 server groups.
- In V100R020C00 and later versions, the DHCPv6 relay service port is disabled by default. Run the **sysman service dhcpv6-relay enable** command to enable the DHCPv6 relay service if necessary.

For the differences of other configuration commands, see "DHCPv6 Configuration" in the *Command Reference*. Some commands need to be executed in diagnose mode. For details about these commands, see "Diagnose Mode Command" in the *Command Reference*.

11.7.6 DHCP Relay Standards and Protocols Compliance

The DHCPv4 relay feature complies with the following standard and protocol:

- RFC 2131: Dynamic Host Configuration Protocol
- DHCPv4 Option 82: RFC 3046

The following lists the reference standards and protocols of the DHCPv6 relay feature:

- DHCPv6 relay: RFC 3315 (DHCPv6 protocol created by IETF)
- DHCPv6 Layer 2 relay: draft-ietf-dhc-dhcpv6-ldra-02 (Draft of a DHCPv6 Layer 2 relay protocol created by IETF), RFC6221
- DHCPv6 option 37: RFC 4649

DHCPv6 option 53: RFC 5460

11.8 DHCP Proxy

This section describes the implementation and configuration of Dynamic Host Configuration Protocol (DHCP) proxy.

11.8.1 What Is DHCP Proxy

Definition

DHCP proxy enables the MA5800 to modify the DHCP packets exchanged between a DHCP server and DHCP clients based on site requirements.

 Copyright © Huawei Technologies Co., Ltd.

DHCP proxy is composed of the server ID proxy and lease proxy functions.

- Server ID proxy

  The Option 54 field in a DHCP packet specifies the IP address of a DHCP server. Server ID proxy enables the MA5800 to change the value of the Option 54 field so that the actual IP address of the DHCP server is hidden from DHCP clients. This reduces the risk of DHCP server attacks initiated by DHCP clients.

- Lease proxy

  The IP address lease that a DHCP client applies for is determined based on Options 51, 58, and 59 in a DHCP packet initiated by the DHCP client. Lease proxy enables the MA5800 to change the values of these options so that the MA5800 provides a shorter lease than that provided by the DHCP server for the DHCP client. This facilitates lease management.

Purpose

The DHCP proxy functions offer solutions for different problems.

- Server ID proxy

  The actual IP address of the DHCP server can be hidden from DHCP clients, which reduces the risk of DHCP server attacks initiated by DHCP clients.

- Lease proxy

  The duration of the lease configured by the DHCP server is so long that the MA5800 cannot obtain the DHCP client status in a timely manner. This obstructs service provisioning.

The lease-time proxy, however, enables a DHCP client to obtain a shorter lease time for an IP address. The MA5800 with the DHCP proxy function enabled is capable of quickly perceiving whether a user is online. Meanwhile, the request packets from the DHCP client for re-leasing an IP address during a short lease time are processed by the MA5800 and are no longer forwarded to the DHCP server. This decreases the load of the DHCP server for frequently processing the request packets during the short lease time.

The MA5800 with DHCP proxy enabled changes the XID values of the DHCP packets sent from DHCP clients to values different from the XID values of the DHCP packets received by the DHCP server. The DHCP server generally does not check XID values. Therefore, the XID value change generally does not affect services. However, if carriers add data to the XID field for data checks on the DHCP server, a data check may fail, which would affect services.

NOTE

Exchange identification (XID) is a field in a DHCP packet that uniquely identifies the DHCP packet.

11.8.2 DHCP Proxy Principles

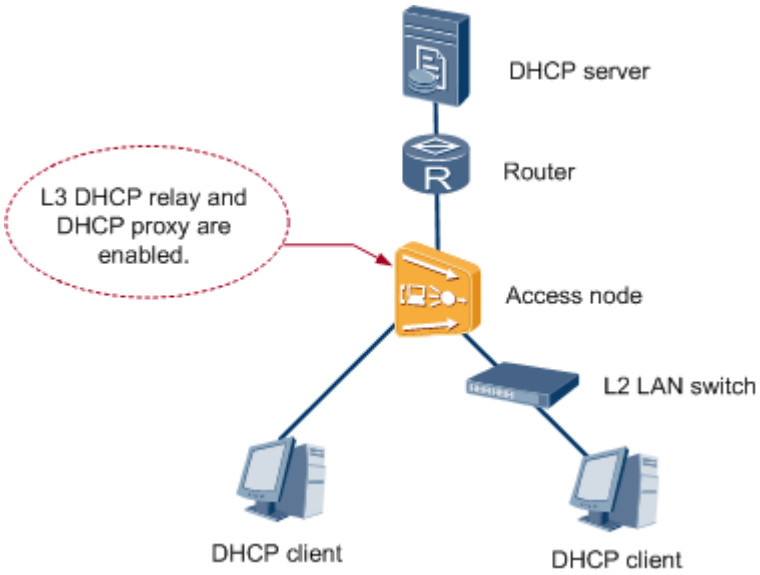
Application Scenario

The MA5800 supports DHCP proxy only when Layer 3 DHCP relay is enabled on it. Both user ports and cascading ports on the MA5800 support DHCP proxy. The MA5800 with DHCP proxy enabled can monitor all DHCP packets exchanged

 Copyright © Huawei Technologies Co., Ltd.

between DHCP clients and a DHCP server. **Figure 11-31** shows an application scenario of DHCP proxy.

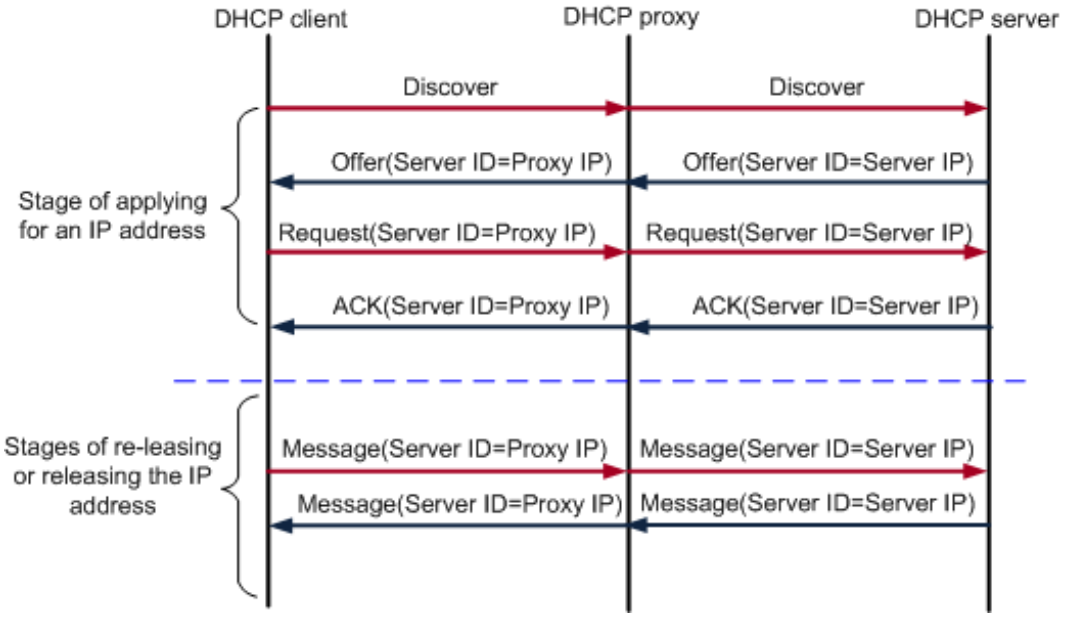
Figure 11-31 Application scenario of DHCP proxy



Server ID Proxy

With DHCP proxy enabled, the server ID proxy function alters the exchange of packets as follows:

Figure 11-32 Exchange of packets between a DHCP client and DHCP server (server ID proxy)



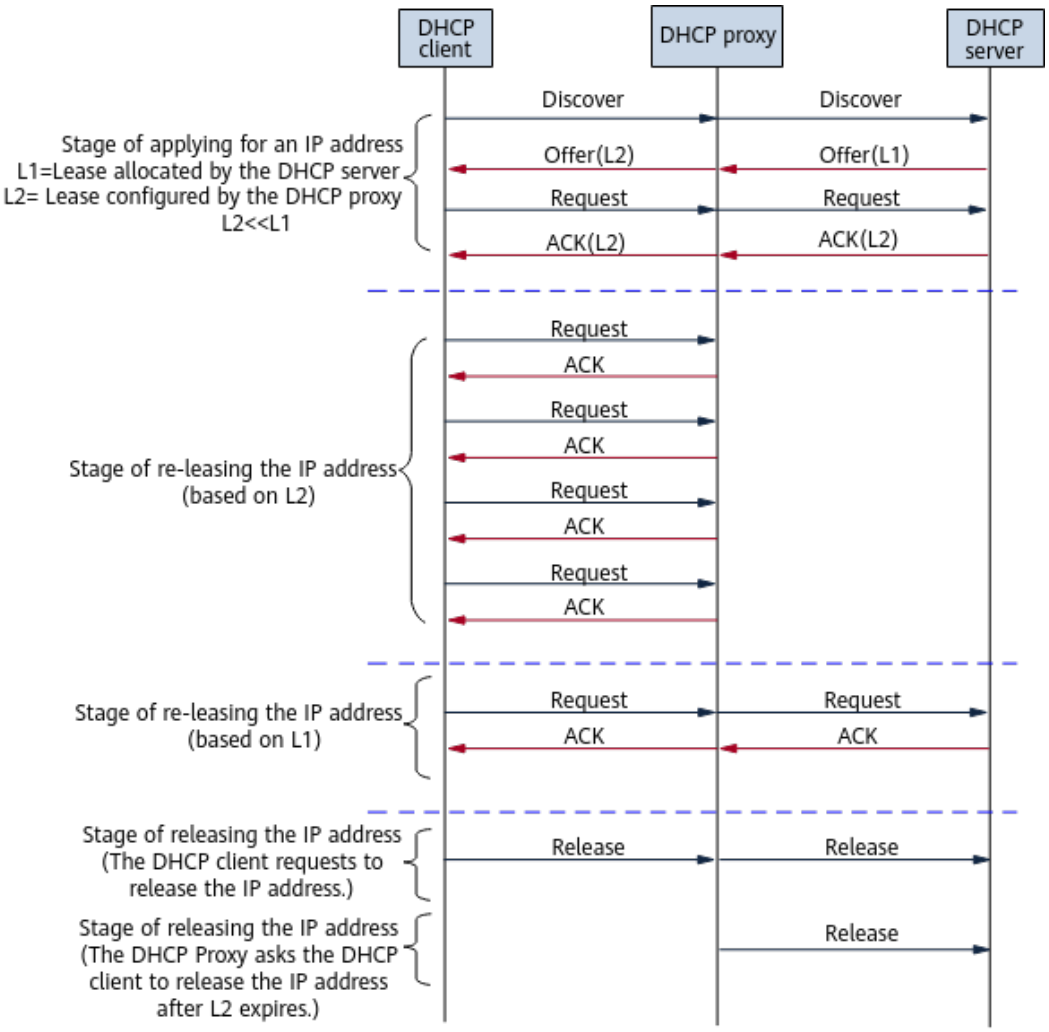
 Copyright © Huawei Technologies Co., Ltd.

- In the downstream direction, the MA5800 changes the value of the Option 54 field in the response packets, such as the Offer and ACK packets, sent by the DHCP server to the IP address of the MA5800. After the DHCP client receives the packets, Option 54 in these packets contains the IP address of the MA5800, and the related field in the DHCP packets is always the IP address of the MA5800 hereafter.
- In the upstream direction, the MA5800 recovers the value of Option 54 in the DHCP packets sent from the DHCP client to the IP address of the actual DHCP server.

Lease Proxy

With DHCP proxy enabled, exchange of packets between a DHCP client and DHCP server (lease proxy) as follows:

Figure 11-33 Exchange of packets between a DHCP client and DHCP server (lease proxy)



Stage of applying for an IP address:

 Copyright © Huawei Technologies Co., Ltd.

1. The DHCP client sends a request packet to the DHCP server for an IP address. The DHCP server then sends a response packet and allocates an IP address with lease L1 to the DHCP client.
2. The MA5800 captures the response packet from the DHCP server, changes the L1 value in the packet to a smaller value L2 (configurable on the MA5800), and sends the Offer packet containing lease L2 to the DHCP client. Then, the lease of the IP address allocated to the DHCP client is changed to L2.

Stage of re-leasing the IP address:

1. The DHCP client sends a request packet to the DHCP server to re-lease the IP address when half of the L2 lease duration has elapsed.
2. The MA5800 captures the request packet and determines whether to send the request packet to the DHCP server based on L1.
   - If the lease time does not reach half of the L1, the MA5800 does not need to send a lease renewal request to the DHCP server, but directly responds to the lease renewal request from the DHCP client, updates the lease renewal time to L2, and allows the client to renew the IP address lease.
   - If the elapsed lease is greater than or equal to half of L1, the MA5800 forwards the request packet to the DHCP server.
3. After receiving the request packet, the DHCP server performs the following operations:
   - Responds to the request packet if it allows the re-leasing. Then, the MA5800 forwards the response packet to the DHCP client.
   - Sends a NACK packet to the DHCP client if it does not allow the re-leasing. Then, the MA5800 notifies the DHCP client that this IP address will be released.

Stage of releasing the IP address:

- If the DHCP client sends a request packet to release the IP address, the MA5800 forwards the request packet to the DHCP server.
- If the MA5800 detects that lease L2 of the DHCP client has expired and the MA5800 has not received any request to re-lease the IP address from the DHCP client, the MA5800 sends a request to release the IP address to the DHCP server.

11.8.3 DHCP Proxy Standards and Protocols Compliance

Reference documents:

- dsl2006[1].127.00: Proposals of DHCP relay improvements

11.9 BFD

11.9.1 Introduction to BFD

The link detection mechanisms of access devices include: slow Hello mechanism, dedicated detection mechanism, and bidirectional forwarding detection (BFD) mechanism.

 Copyright © Huawei Technologies Co., Ltd.

- Slow Hello mechanism: It usually refers to the Hello mechanism offered by a routing protocol. This mechanism can detect a fault in seconds. In high-speed data transmission, for example, at gigabit rates, fault detection longer than 1s will bring the loss of a large amount of data. For delay-sensitive services such as voice services, the delay longer than 1s is also unacceptable.
- Dedicated detection mechanism: Different protocols or manufacturers may provide proprietary detection mechanisms; however, it is difficult to deploy the proprietary detection mechanisms when systems are interconnected.
- BFD detection mechanism: provides uniform detection for all media and protocol layers. It provides light-load and millisecond-level fast fault detection for channels between adjacent forwarding devices. The detected faults may occur on interfaces, data links, or forwarding devices.

In view of the pros and cons of the preceding detection mechanisms, the BFD detection mechanism is widely used in access devices.

11.9.2 Principles to BFD

BFD is used to detect communication link faults between forwarding devices. Specifically, it is used to detect the connectivity of a data protocol on a path between devices. The path can be a physical link, logical link, or tunnel.

- Upper-layer applications provide BFD with parameters, such as the detection address and the detection time.
- BFD creates, deletes, or modifies a BFD session based on these parameters and notifies the upper-layer applications of the session status.

The following describes the basic principles of BFD in terms of the BFD detection mechanism, session establishment mode, and session management.

BFD Detection Mechanism

In the BFD detection mechanism, two systems establish a BFD session and periodically send BFD control packets along the path between them. If one system does not receive BFD control packets within a specified period, the system considers that a fault has occurred on the path, selects another path for communication based on the existing routes, and reports a link session fault alarm.

BFD control packets are encapsulated in UDP packets. In the initial phase of a BFD session, both systems negotiate with each other using parameters in BFD control packets, such as discriminators, expected minimum intervals for sending and receiving BFD control packets, and local BFD session status. After the negotiation succeeds, the two systems transmit BFD control packets to each other at the negotiated interval.

BFD provides the following detection modes:

- Asynchronous mode: most commonly-used BFD detection mode. In this mode, two systems periodically send BFD control packets to each other. If a system fails to receive any packets within a specified period, the BFD session is considered Down.
- Demand mode: an alternative BFD detection mode. A system with a large number of BFD sessions can use the demand mode to prevent overheads

 Copyright © Huawei Technologies Co., Ltd.

generated by periodic BFD control packet transmission from affecting system operation. In this mode, once a BFD session is established, both systems stop sending BFD control packets to each other. Instead, the systems use another mechanism, such as the Hello mechanism of a routing protocol, to check link connectivity. In this manner, the systems are less affected by overheads resulting from BFD sessions.

BFD Session Establishment Modes

A BFD session can be set up in the following modes:

BFD differentiates sessions by My Discriminator and Your Discriminator in the control packets. The main difference in establishment of static and dynamic BFD sessions is that My Discriminator and Your Discriminator are set differently.

Table 11-8 BFD session establishment modes

BFD Session Establishment Mode	Description
Static mode	BFD session parameters, such as the local and remote discriminators, are manually configured and delivered to establish BFD sessions. **NOTE** In static mode, configure unique local and remote discriminators for each BFD session. This mode prevents incorrect discriminators from affecting BFD sessions that have correct discriminators and prevents BFD session flapping.
Dynamic mode	When a BFD session is dynamically established, the system processes the local and remote discriminators as follows: • Dynamically allocates the local discriminator. When a system triggers the dynamic establishment of a BFD session, the system allocates a dynamic discriminator as the local discriminator of the BFD session. Then, the system sends a BFD control packet with Your Discriminator set to 0 to the peer for session negotiation. • Automatically learns the remote discriminator. The local end of a BFD session sends a BFD control packet with **Your Discriminator** set to **0** to the remote end. After the remote end receives the packet, it checks whether the value of **Your Discriminator** in this packet matches the value of its **My Discriminator**. If they are matched, the remote end learns the value of **My Discriminator** of the local end and obtains its **Your Discriminator**.

 Copyright © Huawei Technologies Co., Ltd.

BFD Session Management

A BFD session has the following states:

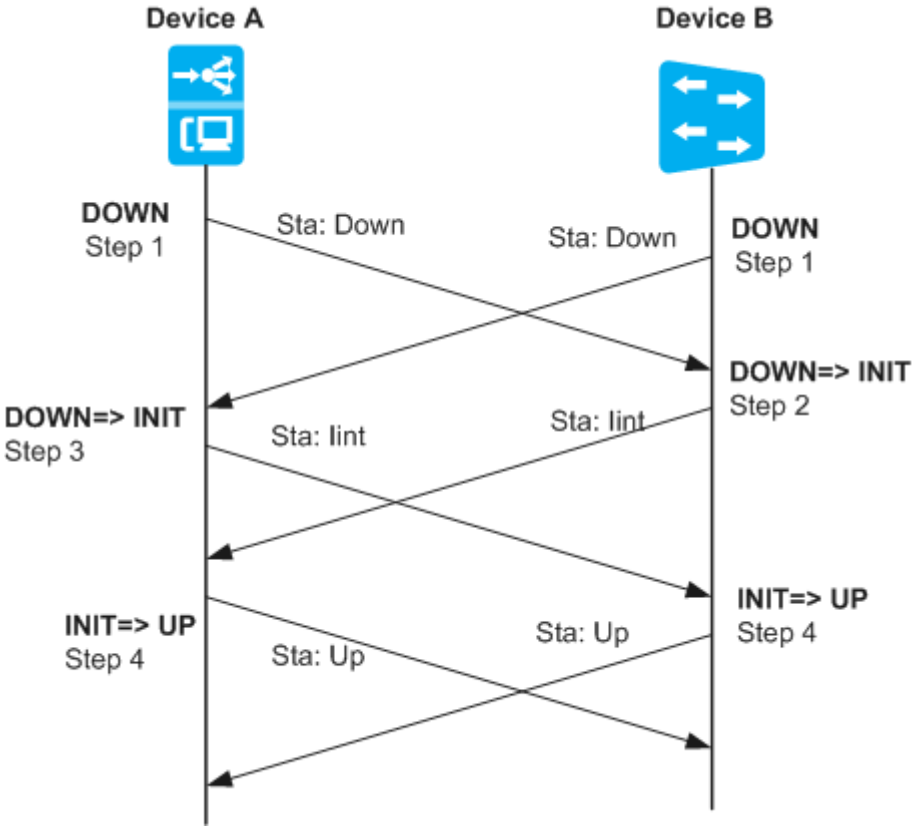
- **Down**: indicates that the BFD session is in the **Down** state or has been set up.
- **Init**: indicates that the local system can communicate with the remote system, and the local system expects a BFD session to go **Up**.
- **Up**: indicates that the BFD session is set up successfully.
- **AdminDown**: indicates that the BFD session is in the **AdminDown** state. After entering the **AdminDown** state, BFD sessions are not deleted but they do not monitor links.

The session status is conveyed in the State field of a BFD control packet. The system changes the session status based on the local session status and the received session status of the peer.

When a BFD session is to be set up or deleted, the BFD state machine implements a three-way handshake to ensure that both two systems are aware of the status change.

As shown in **Figure 11-34**, a BFD session is established as an example to describe the state machine migration process.

Figure 11-34 BFD session establishment



1. Devices A and B enable their BFD session state machines. Because the initial status of the state machines is **Down**, devices A and B send BFD control packets with the **State** field set to **Down**. For a statically established BFD

 Copyright © Huawei Technologies Co., Ltd.

session, the value of **Your Discriminator** in a BFD control packet is specified manually. For a dynamically established BFD session, the value of **Your Discriminator** is **0**.

2. After receiving a BFD control packet with the **State** field set to **Down**, device B changes the status to **Init** and sends BFD control packets with the **State** field set to **Init**.

   NOTE

   After the local BFD session status changes to **Init**, device B no longer processes the received BFD control packets with the **State** field set to **Down**.

3. The BFD status change of device A is the same as that of device B.
4. After receiving a BFD control packet with the **State** field set to **Init**, device B changes the local BFD session status to **Up**.
5. The BFD status change of device A is the same as that of device B.

11.9.3 BFD Echo

BFD echo is a rapid fault detection mechanism in which the local system sends BFD echo packets and the remote system loops back the packets. BFD echo is classified into passive BFD echo and one-arm BFD echo modes. These two BFD echo modes have the same detection mechanism but different application scenarios.

Passive BFD Echo

The device supports passive BFD echo for interworking with other vendors' devices.

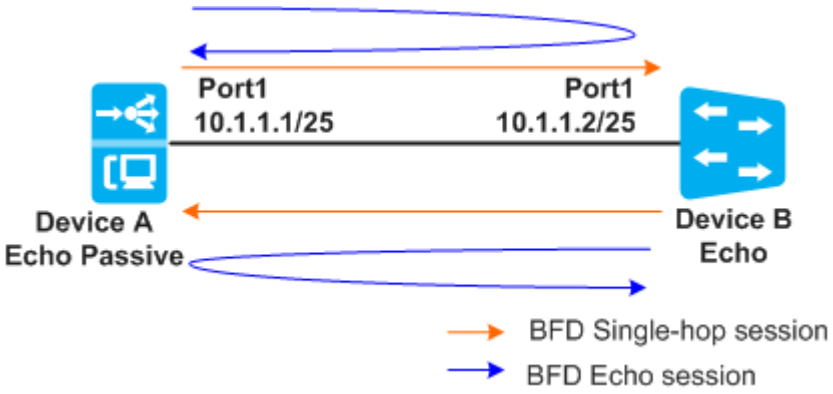
Passive BFD echo applies only to single-hop IP link scenarios and works with asynchronous BFD. When a BFD session works in asynchronous echo mode, the two endpoints of the BFD session perform both slow detection in asynchronous mode and quick detection in echo mode.

As shown in **Figure 11-35**, Device A is directly connected to Device B, and asynchronous BFD sessions are established between the two devices. After active BFD echo is enabled on Device B and passive BFD echo is enabled on Device A, the two devices work in asynchronous echo mode and send single-hop and echo packets to each other.

If Device A has a higher BFD performance than Device B, for example, the minimum intervals between receiving BFD packets supported by Device A and Device B are 3 ms and 100 ms respectively, then BFD sessions in asynchronous mode will adopt the larger interval (100 ms). If BFD echo is enabled, Device A can use echo packets to implement faster link failure detection. If BFD echo is disabled, Device A and Device B can still use asynchronous BFD packets to detect link failures. However, the minimum interval between receiving BFD packets is the larger interval value (100 ms in this example).

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-35 Passive BFD echo networking



The process of establishing a passive BFD echo session as shown in **Figure 11-35** is as follows:

1. Device B functions as a BFD session initiator and sends an asynchronous BFD packet to Device A. The Required Min Echo RX Interval field carried in the packet is a nonzero value, which specifies that Device A must support BFD echo.
2. After receiving the packet, Device A finds that the value of the Required Min Echo RX Interval field carried in the packet is a nonzero value. If Device A has passive BFD echo enabled, it checks whether any ACL that restricts passive BFD echo is referenced. If an ACL is referenced, only BFD sessions that match specific ACL rules can enter the asynchronous echo mode. If no ACL is referenced, BFD sessions immediately enter the asynchronous echo mode.
3. Device B periodically sends BFD echo packets, and Device A sends BFD echo packets (the source and destination IP addresses are the local IP address, and the destination physical address is Device B's physical address) at the interval specified by the Required Min RX Interval field. Both Device A and Device B start a receive timer, with a receive interval that is the same as the interval at which they each send BFD echo packets.
4. After Device A and Device B receive BFD echo packets from each other, they immediately loop back the packets at the forwarding layer. Device A and Device B also send asynchronous BFD packets to each other at an interval that is much less than that for sending echo packets.

One-Arm BFD Echo

One-arm BFD echo applies only to single-hop IP link scenarios. Generally, one-arm BFD echo is used when two devices are directly connected and only one of them supports BFD. Therefore, one-arm BFD echo does not require both ends to negotiate echo capabilities. A one-arm BFD echo session can be established on a device that supports BFD. After receiving a one-arm BFD echo session packet, devices that do not support BFD immediately loop back the packet, implementing quick link failure detection.

The local device that has one-arm BFD echo enabled sends a special BFD packet (both the source and destination IP addresses in the IP header are the local IP address, and the MD and YD in the BFD payload are the same). After receiving the packet, the remote device immediately loops the packet back to the local device

 Copyright © Huawei Technologies Co., Ltd.

to determine link reachability. One-arm BFD echo can be used on low-end devices that do not support BFD.

Similarities and Differences Between Passive BFD Echo and One-Arm BFD Echo

To ensure that passive BFD echo or one-arm BFD echo can take effect, disable strict URPF on devices that send BFD echo packets.

Strict URPF prevents attacks that use spoofed source IP addresses. If strict URPF is enabled on a device, the device obtains the source IP address and inbound interface of a packet and searches the forwarding table for an entry with the destination IP address set to the source IP address of the packet. The device then checks whether the outbound interface for the entry matches the inbound interface. If they do not match, the device considers the source IP address invalid and discards the packet. After a device enabled with strict URPF receives a BFD echo packet that is looped back, it checks the source IP address of the packet. As the source IP address of the echo packet is a local IP address of the device, the packet is sent to the platform without being forwarded at the lower layer. As a result, the device considers the packet invalid and discards it.

Table 11-9 Differences between BFD echo sessions and common static single-hop sessions

BFD Session	Supported IP Type	Session Type	Descriptor	Negotiation Prerequisite	IP Header
Common static single-hop session	IPv4 and IPv6	Static single-hop session	MD and YD must be configured.	A matching session must be established on the peer.	The source and destination IP addresses are different.
Passive BFD echo session	IPv4 and IPv6	Dynamic single-hop session	No MD or YD needs to be configured.	A matching session must be established and echo must be enabled on the peer.	Both the source and destination IP addresses are a local IP address of the device.
One-arm BFD echo session	IPv4	Static single-hop session	Only MD needs to be configured (MD and YD are the same).	A matching session does not need to be established on the peer.	Both the source and destination IP addresses are a local IP address of the device.

11.9.4 Application Environment

 Copyright © Huawei Technologies Co., Ltd.

11.9.4.1 BFD for IP

A BFD session is established on an IP link to fast detect faults.

BFD can detect single-hop and multi-hop IP links.

- Single-hop BFD detects IP route connectivity between directly-connected systems. The single hop refers to an IP hop. Between these two systems, only one BFD session can be set up for a specified data protocol on an interface.
- Multi-hop BFD detects any paths between systems. A path may span multiple hops or may partially overlap.

Example 1

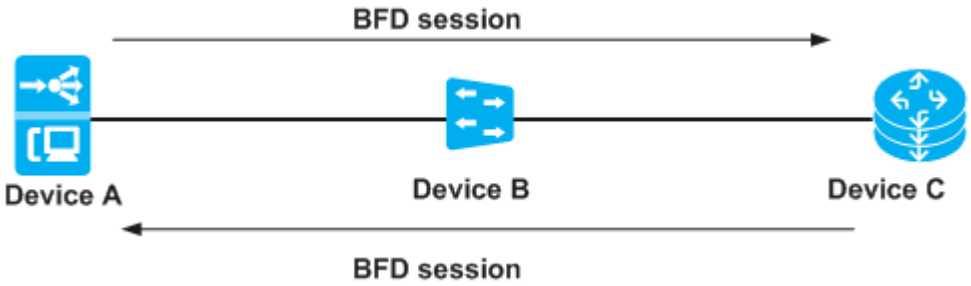
Figure 11-36 shows a single-hop BFD session detecting a path between directly-connected devices. The BFD session is bound to the outgoing interface.

Figure 11-36 Single-hop BFD for IP

Example 2

Figure 11-37 shows a multi-hop BFD session detecting a path between Device A and Device C. The BFD session is bound to the peer IP address but not the outgoing interface.

Figure 11-37 Multi-hop BFD for IP



11.9.4.2 BFD for OSPF

A link fault or change in topology may lead to rerouting in a network. Quick convergence of a routing protocol is important for improving network availability.

A feasible solution is to fast detect the fault and immediately notify the routing protocol of the fault.

 Copyright © Huawei Technologies Co., Ltd.

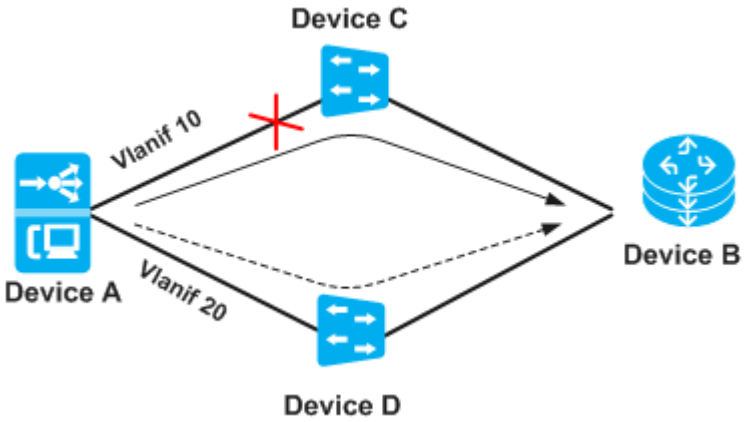
In BFD for OSPF, OSPF is associated with a BFD session. The BFD session fast detects a link fault and notifies OSPF of the fault. In this manner, OSPF speeds up responses to changes in network topology.

Table 11-10 shows convergence speed statistics when OSPF is and is not associated with a BFD session.

Table 11-10 OSPF convergence speed statistics

Associated with BFD	Link Fault Detection Mechanism	Convergence Speed
No	OSPF Hello keepalive timer timeout	Within seconds
Yes	BFD session in the Down state	Within milliseconds

Figure 11-38 BFD for OSPF networking diagram



As shown in Figure 11-38, Device A sets up OSPF neighbor relationships with Device C and Device D. The outbound interface VLANIF 10 on Device A is connected to Device B through Device C. When the neighbor state is Full, BFD is notified of the status and starts to set up a BFD session.

1. When a fault occurs on the link between Device A and Device C, the BFD session detects the fault and notifies Device A.
2. Device A processes the neighbor-Down event and recalculates routes. Then, the outbound interface changes to VLANIF 20 on Device A, which is connected Device B through Device D.

11.9.4.3 BFD for IS-IS

Generally, the interval at which the Intermediate System to Intermediate System (IS-IS) protocol sends Hello messages is 10 seconds. If a device does not receive any Hello message from its neighbor within three Hello intervals, the device deletes the neighbor. Therefore, it takes a device a number of seconds to detect that a neighbor is Down. This leads to the loss of a large number of packets in a high-speed network.

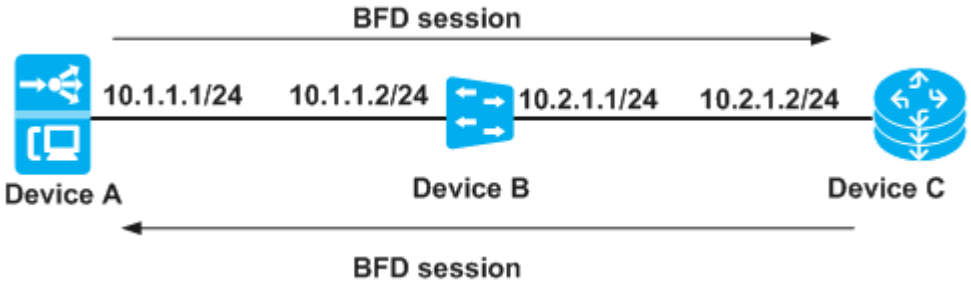
 Copyright © Huawei Technologies Co., Ltd.

In BFD for IS-IS, the establishment of a BFD session is dynamically triggered by IS-IS but not configured manually. When detecting a fault, the BFD session notifies IS-IS of the fault through the Routing Management Module (RM). IS-IS processes the neighbor-Down event and quickly sends the link state PDU (LSP), and performs the partial route calculation (PRC). In this manner, IS-IS routes fast converge.

The BFD fault detection interval is at the millisecond level. Instead of replacing the IS-IS Hello mechanism, BFD works with IS-IS to detect the adjacency fault more quickly. In addition, BFD instructs IS-IS to recalculate routes, ensuring correct packet forwarding.

The RM allows IS-IS and BFD to interact with each other. Through the RM, IS-IS instructs BFD to dynamically set up or delete BFD sessions. The BFD event messages are also delivered to IS-IS through the RM.

Figure 11-39 BFD for IS-IS networking diagram



After BFD is enabled on Device A, Device B, and Device C, the BFD session can quickly detect faults on the link between Device A and Device B, and notify IS-IS through the RM. Then, IS-IS sets the neighbor status to Down to trigger the IS-IS topology calculation. In addition, IS-IS updates LSPs to ensure that Device C (Device B's neighbor) can receive the updated LSPs from Device B in time. This implements fast network topology convergence.

11.9.4.4 BFD for BGP

The Border Gateway Protocol (BGP) periodically sends Keepalive messages to its peer to monitor the neighbor status. This detection process lasts more than 1 second. When the data is transmitted at gigabit rates, a large amount of data will be discarded, which cannot meet the requirement for carrier-class reliability.

BFD for BGP was developed to compensate for this shortcoming. The BFD session can fast detect a fault on a link between BGP peers and notify BGP, ensuring fast convergence.

NOTE

By default, a multi-hop BGP session is established between Huawei devices that set up an IBGP peer relationship. A BFD for IGP session and A BFD for IBGP session cannot be both set up between a Huawei device and a non-Huawei device that sets up a single-hop BGP session with its peer by default. In such a situation, setting up only A BFD for IGP session or A BFD for IBGP session between the Huawei and non-Huawei devices is recommended.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-40 BFD for BGP networking

As shown in **Figure 11-40**, Device A and Device B are directly connected through the External Border Gateway Protocol (EGBP). A BFD session is established to detect the BGP neighbor relationship between Device A and Device B. When the link between Device A and Device B is faulty, the BFD session can quickly detect the fault and notify BGP.

11.9.4.5 BFD for RSVP

BFD monitors RSVP neighbor relationships. When a Layer 2 device (a hub for example) exists between RSVP neighboring nodes, the two nodes can detect a link fault only using the Hello mechanism in seconds. This process results in the loss of lots of data. BFD for RSVP rapidly detects faults in a link between RSVP neighboring nodes within milliseconds. BFD for RSVP applies to TE FRR networks, on which Layer 2 devices exist on a primary CR-LSP between the PLR and its RSVP neighboring node.

BFD for RSVP is an IP-layer detection. Only a single-hop BFD session can be set up between RSVP neighboring nodes.

Figure 11-41 BFD for RSVP

A BFD session for RSVP is set up to monitor the link between RSVP neighbors. The RSVP module can rapidly detect a link failure.

BFD for RSVP can share BFD sessions with BFD for Open Shortest Path First (OSPF), BFD for Intermediate System to Intermediate System (IS-IS), or BFD for Border Gateway Protocol (BGP). The local node selects the smallest values of parameters between the two ends of the shared BFD session as local BFD parameters. The parameters include the interval at which BFD packets are sent, interval at which BFD packets are received, and local detection multiplier.

 Copyright © Huawei Technologies Co., Ltd.

11.9.4.6 BFD for LAG

Link aggregation group (LAG) is a feature to achieve load balancing and protection of Ethernet links. By bundling multiple Ethernet links between 2 Ethernet switching devices into a LAG, the Ethernet links can form a high-bandwidth Ethernet link. Ethernet traffic between the 2 Ethernet switching devices is automatically distributed on member links to implement load balancing. When a member link becomes faulty, services are automatically distributed to other normal links to implement protection switching.

The LAG feature detects faults on member links by using either of the following mechanisms:

- Port status detection. A port in the **Down** state is determined faulty. The detection is fast, but only physical-layer faults can be detected. In some Ethernet working modes, the faults of only the receive links can be detected.
- LACP protocol detection. Ethernet switching devices at both ends of a LAG periodically send LACP frames to each other. By detecting LACP frames and parsing the information carried in the LACP frames, the Ethernet switching devices detect whether a link is faulty. The detection is comprehensive, but the LACP protocol is a slow protocol and the sending interval is long (in seconds). Therefore, the detection is slow.

Figure 11-42 BFD for LAG

Based on the exchange of BFD packets, the BFD for LAG feature can quickly detect various forwarding faults of LAG member links and distribute the services on the faulty link to other normal links in a timely manner. This overcomes the limitations of the 2 detection mechanisms of the LAG feature. The BFD for LAG feature has the following advantages:

- BFD for LAG has a faster detection speed than LACP and can substitute for LACP.
- BFD for LAG can detect Layer 3 packet forwarding capabilities. In contrast, LACP is a Layer 2 protocol and can detect only Layer 2 forwarding capabilities.

11.9.4.7 BFD for PW

The bidirectional forwarding detection (BFD) for pseudo wire (PW) is used to quickly check the connectivity of the PWs in virtual private wire service (VPWS) and virtual private LAN service (VPLS) applications, providing a basis for fast switching for PW redundance. BFD is a fast detection mechanism, which consumes a lot of resources. Generally, BFD is used only for key PWs.

 Copyright © Huawei Technologies Co., Ltd.

Implementation Principle

In BFD for PW detection, BFD packets are encapsulated in a PW and sent periodically to check the connectivity of the PW.

This version supports the following types of BFD for PW:

- Static BFD for VPWS
- Dynamic BFD for VPWS
- Static BFD for VPLS

NOTE

Static sessions are implemented by Huawei. To interconnect with devices of other vendors, you can use dynamic BFD sessions.

In the PW standard, VCCV is used to define the connectivity check channel. That is, continuity check (CC) and CV are used to define the channel type and check type for connectivity check.

- **CC type:**
  - Control word mode: This field is encapsulated to indicate that a packet is a BFD detection packet.
  - Router Alert Label mode: This field is encapsulated to indicate that a packet is a BFD detection packet.
  - TTL mode: The TTL expiration mode is used to indicate that a packet is a BFD detection packet.

  This version supports only the control word mode.
- **CV type:**

Bit	**Value**	**MPLS PW CV Type**
Bit 0	**0x01**	**ICMP Ping**
Bit 1	**0x02**	**LSP Ping**
Bit 2	**0x04**	Indicates IP or UDP BFD.
Bit 3	**0x08**	Indicates IP/UDP BFD +status notification.
Bit 4	**0x10**	Indicates raw BFD.
Bit 5	**0x20**	Indicates raw BFD +status notification.
Bit 6	**0x40**	Reserved.
Bit 7	**0x80**	Reserved.

The BFD feature of this version supports the following CV types:

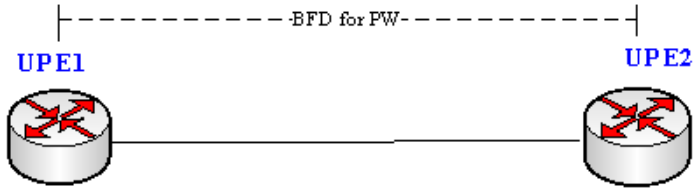
- All supported BFD for PW session types support 0x04.

 Copyright © Huawei Technologies Co., Ltd.

– Dynamic BFD for VPWS also supports 0x10.

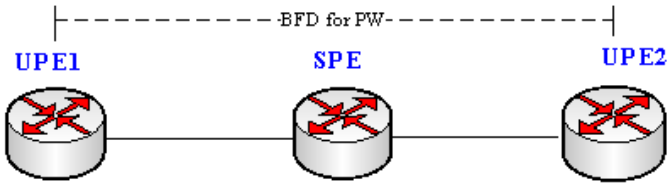
VPWS Scenario

Figure 11-43 Checking the connectivity between PEs for a single-hop PW



In this scenario, the PW connectivity between UPE1 and UPE2 is checked.

Figure 11-44 Checking the connectivity between PEs for a multi-hop PW.



In this VPWS scenario, the superstratum provider edge (SPE) in the middle switches PW labels. This version does not support the PW connectivity check between UPE1 and the SPE.

NOTE

In this scenario, the MA5800 cannot function as an SPE.

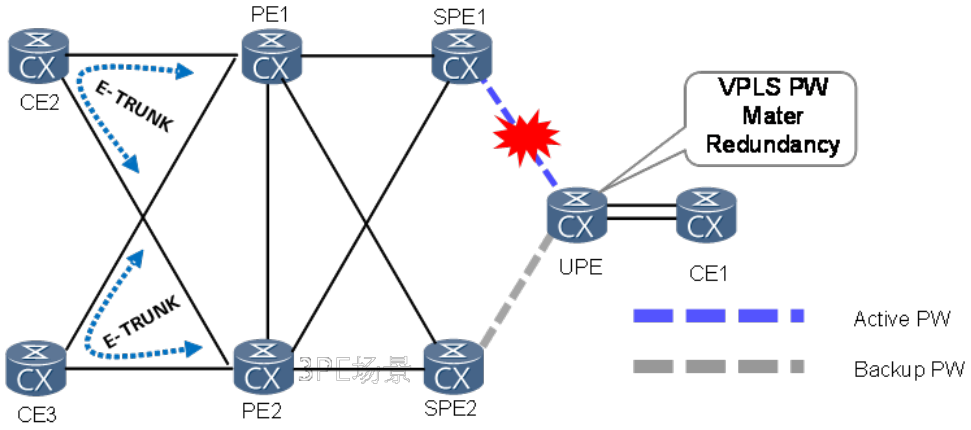
VPLS Scenario

In VPLS scenarios, BFD for PW cannot traverse SPEs on the hierarchical virtual private LAN service (HVPLS) network. That is, BFD for PW can only detect the segment from a UPE to an SPE on the VPLS network.

 Copyright © Huawei Technologies Co., Ltd.

As shown in the figure, the customer edge (CE) is dual-homed to the VPLS network, and Eth-Trunk is deployed on the CE and PE to ensure link reliability. The CE communicates with remote devices on the VPLS network through PE1.

Figure 11-45 Failure of the primary PW between the UPE and SPE1



VPLS PW redundancy is configured between SPEs and UPEs, and BFD is configured for the primary and secondary PWs.

Failover: Upon detecting that the primary PW is down, the UPE switches traffic to the secondary PW and sends MAC Withdraw messages in which the PE ID field carries the SPE1 LSR ID to SPE2. SPE2 transparently transmits the MAC Withdraw messages to PE1 and PE2. SPE2, PE1, and PE2 clear the MAC addresses learned from SPE1.

Switchback: After the primary PW recovers, the BFD negotiation is up, and the UPE instructs SPE2 to change its PW forwarding status to standby and SPE1 to change its PW forwarding status to active. The UPE sends MAC Withdraw messages in which the PE ID field carries the SPE2 LSR ID to SPE1. SPE1 transparently transmits the MAC Withdraw messages to PE1 and PE2. SPE1, PE1, and PE2 clear the MAC addresses learned from SPE2. PE1 and PE2 then relearn MAC addresses by broadcasting upon receiving multicast packets from the primary PW.

11.9.5 Configuring the BFD

This topic describes how to configure the BFD on the MA5800.

Configuration Scenario Description

- When you need to configure BFD in IP static route, OSPF, IS-IS, BGP, RSVP, and LAG scenarios, refer to this section to simplify the configuration process.
- When you need to configure links not belonging to the preceding scenarios, configure the links one by one by referring to "Configuring BFD Sessions" in this chapter. If BFD echo is required, refer to **11.9.5.2 Configuring One-Arm BFD Echo Sessions** to make the configurations.

11.9.5.1 Configuring BFD Sessions

A bidirectional forwarding detection (BFD) session rapidly detects faults in links over a network.

 Copyright © Huawei Technologies Co., Ltd.

Context

In the BFD detection mechanism, two systems set up a BFD session, and periodically send BFD control packets along the path between them. If one system does not receive BFD control packets within a specified period, the system considers that a fault occurs on the path.

BFD uses the local and remote discriminators to differentiate multiple BFD sessions between the same pair of systems. Based on the differences in methods of creating the local and the remote discriminators, MA5800 supports the following types of BFD sessions:

- Static BFD sessions with manually-specified discriminators

  The local and remote discriminators must be set manually. The discriminators on the remote end must also be manually specified.
- Static BFD sessions with automatically-negotiated discriminators

  If a dynamic BFD session is used by a remote device, a static BFD session with automatically negotiated discriminators must be created on a local device to interwork with the remote device and support the BFD for static routes. The discriminators on the remote end can be automatically negotiated or a dynamic BFD session can be established on the remote end.
- For BFD sessions dynamically triggered by protocols, no local or remote discriminator needs to be set. The implementation method is as follows:
  - BFD sessions with dynamically-allocated local discriminators.
  - BFD sessions with self-learned remote discriminators.

Procedure

Step 1 Enable BFD globally.

1. Run the **bfd** command to enable BFD globally and enter the BFD mode.

   BFD must be enabled globally before configurations relevant to BFD are performed. By default, BFD is disabled globally.
2. Run the **quit** command to quit the BFD mode.

Step 2 Create a BFD session. Select one of the following steps depending on the type of link to be checked by BFD.

- For an IPv4 link

  Run the **bfd bind peer-ip** command to create a BFD session.

  If the **bfd bind peer-ip source-ip auto** command is run, a BFD session is set up through automatic negotiations over discriminators. The device on which such a BFD session is created can interoperate with another device on which a dynamic BFD is set up. This command is used to configure BFD sessions for IPv4 static routes.
- For an IPv6 link

  Run the **bfd bind peer-ipv6** command to create a BFD session.

  If the **bfd bind peer-ipv6 source-ipv6 auto** command is run, a BFD session is set up through automatic negotiations over discriminators. The device on which such a BFD session is created can interoperate with another device on which a dynamic BFD is set up. This command is used to configure BFD sessions for IPv6 static routes.

 Copyright © Huawei Technologies Co., Ltd.

Pay attention to the following points:

- If a single-hop BFD session is to be set up on an interface for the first time, the interface and its peer address must be bound to the BFD session. The bindings cannot be modified after the BFD session is successfully created.
- If a multi-hop BFD session is to be set up on an interface for the first time, the peer address must be bound to the BFD session. The bindings cannot be modified after the BFD session is successfully created.
- During BFD configuration items are being created, the system checks only the format, not the correctness, of an IP address. Either an incorrect peer or source IP address leads to a failure in creating a BFD session.

Step 3 Configure the discriminators.

1. Run the **discriminator local** *discr-value* command to configure a local discriminator.
2. Run the **discriminator remote** *discr-value* command to configure a remote discriminator.

The local discriminator set on a device is equal to the remote discriminator set on a remote device, and the remote discriminator set on the local device is equal to the local discriminator set on the remote device. If the discriminators on the device and the remote device do not match, the session cannot be created. After the local and remote discriminators are set, they cannot be changed.

Step 4 (Optional) Configure the BFD parameters.

Select the following desired operations:

- Modify the detection time.
  - Run the **min-tx-interval** command to configure the interval for sending BFD packets.
  - Run the **min-rx-interval** command to configure the interval for receiving BFD packets.
- Run the **detect-multiplier** command to configure the local detection multiplier.
- Run the **description** command to add the description of a BFD session. Descriptions of BFD sessions help you distinguish between various BFD sessions.

  The **description** command takes effect only on the statically configured BFD sessions, rather than the BFD sessions that are dynamically configured or the BFD sessions that are set up through automatic negotiations over discriminators.
- Run the **tos-exp** command to configure the priority of the BFD packet. By default, the highest priority 7 is adopted. When the system is congested, the BFD packet with higher priority can be sent first.

  The command is valid in static BFD mode. The command is invalid in dynamic BFD mode.
- Run the **wtr** command to configure the time of waiting for recovery of the BFD session. By default, the value is 0, indicating no waiting.

  The BFD session is unidirectional. The detection is performed by BFD parameters configured on both ends respectively. If wait-to-recovery (WTR) is

 Copyright © Huawei Technologies Co., Ltd.

needed, configure it on two ends manually. Or, when the status of the session on one end changes, the applications on both ends can find that the states of the BFD sessions are inconsistent.

Step 5 Query the BFD session information and BFD session statistics.

- Run the **display bfd configuration** command to query the BFD configuration.
- Run the **display bfd interface** command to query the BFD configuration on an interface.
- Run the **display bfd session** command to query the BFD session information.
- Run the **display bfd statistics** command to query the BFD global statistics.
- Run the **display bfd statistics session** command to query the BFD session statistics.

----End

Example

Example: VLAN 10 has been created, and the IP address of VLANIF 10 has been configured. Single-hop BFD needs to be configured on VLANIF 10. Assume that the peer IP address is 10.1.1.1/24, BFD session name is test, the local discriminator is 100, the remote discriminator is 200, the minimum transmit interval and minimum receive interval of BFD control packets are both 10 milliseconds, the local detection multiplier is 3 (default value). The OLT is configured with xPON dual-homing protection.

```
Configure a BFD session as follows:
huawei(config)#bfd
huawei(config-bfd)#quit
huawei(config)#bfd test bind peer-ip 10.1.1.1 interface vlanif 10
huawei(config-bfd-session-test)#discriminator local 100   //The discriminator corresponds to the peer end.
huawei(config-bfd-session-test)#discriminator remote 200   //The discriminator corresponds to the peer end.
huawei(config-bfd-session-test)#min-tx-interval 10
huawei(config-bfd-session-test)#min-rx-interval 10
Add a protection group to the OLT as follows:
huawei(config)#protect-group protect-target gpon-uni-ont workmode dual-parenting
Bind the OLT protection group to the uplink interface, and create an association between the protection group and the BFD session as follows:
huawei(protect-group-0)#uplink-monitor port bfd test   //test is the name configured for the BFD.
Add a member to the OLT protection group and configure a work-side ONT as follows:
huawei(protect-group-0)#protect-group member port 0/1/0 ont 0 role work
```

11.9.5.2 Configuring One-Arm BFD Echo Sessions

A one-arm BFD echo session can be established on a device that supports BFD. After receiving a one-arm BFD echo session packet, devices that do not support BFD immediately loop back the packet, implementing quick link failure detection.

Procedure

Step 1 Enable BFD globally.

1. Run the **bfd** command to enable BFD globally and enter the BFD mode.

   BFD must be enabled globally before configurations relevant to BFD are performed. By default, BFD is disabled globally.

2. Run the **quit** command to quit the BFD mode.

 Copyright © Huawei Technologies Co., Ltd.

Step 2 Run the **bfd one-arm-echo** command to create a one-arm BFD echo session.

Pay attention to the following points:

- If a single-hop BFD session is to be set up on an interface for the first time, the interface and its peer address must be bound to the BFD session. The bindings cannot be modified after the BFD session is successfully created.
- During BFD configuration items are being created, the system checks only the format, not the correctness, of an IP address. Either an incorrect peer or source IP address leads to a failure in creating a BFD session.

Step 3 Run the **discriminator local** *discr-value* command to configure a local discriminator.

•When creating a one-arm BFD echo session, you need to specify only the **local** *discr-value* parameter.

Step 4 (Optional) Configure the BFD parameters.

Select the following desired operations:

- Run the **min-echo-rx-interval** command to configure the interval for receiving BFD packets.
- Run the **detect-multiplier** command to configure the local detection multiplier.
- Run the **description** command to add the description of a BFD session. Descriptions of BFD sessions help you distinguish between various BFD sessions.

  The **description** command takes effect only on the statically configured BFD sessions, rather than the BFD sessions that are dynamically configured or the BFD sessions that are set up through automatic negotiations over discriminators.
- Run the **tos-exp** command to configure the priority of the BFD packet. By default, the highest priority 7 is adopted. When the system is congested, the BFD packet with higher priority can be sent first.
- Run the **wtr** command to configure the time of waiting for recovery of the BFD session. By default, the value is 0, indicating no waiting.

Step 5 Query the BFD session information and BFD session statistics.

- Run the **display bfd configuration** command to query the BFD configuration.
- Run the **display bfd interface** command to query the BFD configuration on an interface.
- Run the **display bfd session** command to query the BFD session information.
- Run the **display bfd statistics** command to query the BFD global statistics.
- Run the **display bfd statistics session** command to query the BFD session statistics.

----End

Example

Assume that the peer IP address is 10.1.1.1/24, BFD session name is test, the local discriminator is 100, the minimum transmit interval and minimum receive interval

 Copyright © Huawei Technologies Co., Ltd.

of BFD control packets are both 10 milliseconds, the local detection multiplier is 3 (default value), VLAN 10 is created, and the IP address of VLAN interface 10 is configured. To configure one-arm BFD echo single-hop detection on VLAN interface 10, run the following commands:

```
huawei(config)#bfd
huawei(config-bfd)#quit
huawei(config)#bfd test bind peer-ip 10.1.1.1 interface vlanif 10 one-arm-echo
huawei(config-bfd-session-test)#discriminator local 100
huawei(config-bfd-session-test)#min-echo-rx-interval 10
```

11.9.5.3 Configuring BFD for IPv4 Static Routes

BFD for IPv4 static routes enables devices to fast detect link changes, improving network reliability.

Prerequisites

Configure parameters of the link layer protocol and IP addresses for interfaces and ensure that the link layer protocol on the interfaces is Up.

Context

Optimal IPv4 static routes are delivered to the forwarding table for packet forwarding. However, IPv4 static routes cannot detect the status of the link to the next hop. You can bind IPv4 static routes to BFD sessions. A BFD session can fast detect changes over a link and inform the routing management system of the changes. The routing management system immediately deletes the IPv4 static route that is bound to the BFD session from the forwarding table and recalculates another active route. In this manner, fast route convergence is implemented.

Procedure

Step 1 Enable BFD globally.

1. Run the **bfd** command to enable BFD globally and enter the BFD mode.
2. Run the **quit** command to quit the BFD mode.

Step 2 Create a static BFD session.

1. Run the **bfd bind peer-ip** command to create a BFD session.
2. Configure the discriminators.

   The local discriminator set on a device is equal to the remote discriminator set on a remote device, and the remote discriminator set on the local device is equal to the local discriminator set on the remote device. If the discriminators on the device and the remote device do not match, the session cannot be created.

   a. Run the **discriminator local** *discr-value* command to configure a local discriminator.
   b. Run the **discriminator remote** *discr-value* command to configure a remote discriminator.
3. Run the **quit** command to quit the BFD session mode.
4. Run the **ip route-static** *ip-address* { *mask* | *mask-length* } *interface-type interface-number* [*nexthop-address*] [**preference** *preference*] **track bfd-**

 Copyright © Huawei Technologies Co., Ltd.

session *cfg-name* [**description** *text*] command to bind a BFD session to an IPv4 static route on the public network.

Step 3 Create a dynamic BFD session.

1. Run the **ip route-static** *ip-address* { *mask* | *mask-length* } *interface-type interface-number* [*nexthop-address*] [**preference** *preference*] **bfd enable** [**description** *text*] command to bind a BFD session to an IPv4 static route on the public network.
2. (Optional) Run the **ip route-static default-bfd** command to configure global BFD parameters for IPv4 static routes.
3. Run the **ip route-static bfd** command to configure BFD parameters for an IPv4 static route.

Step 4 Query the BFD session information.

Run the **display bfd session** command to query the BFD session.

----End

11.9.5.4 Configuring BFD for OSPF

Configuring BFD for OSPF helps the system to rapidly detect the link status and speeds up OSPF convergence in the case of a link failure.

Prerequisites

Basic OSPF functions are configured.

Context

After BFD for OSPF is configured, when detecting a link fault, BFD rapidly notifies the routers on both ends of the link of the fault, triggering rapid OSPF convergence. When the OSPF neighbor relationship goes Down, the BFD session will be dynamically deleted.

Procedure

Step 1 Enable BFD globally.

1. Run the **bfd** command to enable BFD globally and enter the BFD mode.
2. Run the **quit** command to quit the BFD mode.

Step 2 Create a BFD session.

- To configure BFD for all interfaces in a specified OSPF process, perform the following steps:

1. Run the **ospf** command to enter the OSPF mode.
2. Run the **bfd all-interfaces enable** command to enable BFD for all the interfaces in a specific OSPF process. The default parameter values are used.
3. (Optional) Run the **bfd all-interfaces** { **min-rx-interval** *receive-interval* | **min-tx-interval** *transmit-interval* | **detect-multiplier** *multiplier-value* } * command to modify the parameters of BFD session.
4. Run the **quit** command to quit the OSPF mode.

 Copyright © Huawei Technologies Co., Ltd.

5. (Optional) Prevent an interface from dynamically creating a BFD session.

   After BFD for OSPF is configured, all interfaces on which neighbor relationships are Full in the OSPF process will create BFD sessions. To prevent specific interfaces from being enabled with BFD, disable these interfaces from dynamically creating BFD sessions.

   a. Run the **interface vlanif** command to enter the VLANIF interface mode.
   b. Run the **ospf bfd block** command to prevent the interface from dynamically creating a BFD session.
   c. Run the **quit** command to quit the VLANIF interface mode.

- To configure BFD for specified OSPF interfaces, perform the following steps:

1. Run the **interface vlanif** command to enter the VLANIF interface mode.
2. Run the **ospf bfd enable** command to enable BFD for an OSPF interface.
3. (Optional) Run the **ospf bfd** { **min-rx-interval** *receive-interval* | **min-tx-interval** *transmit-interval* | **detect-multiplier** *multiplier-value* } * command to modify the parameters of BFD session.
4. Run the **quit** command to quit the VLANIF interface mode.

Step 3 Query the BFD session information.

Run the **display ospf bfd session** command to query the BFD session.

----End

11.9.5.5 Configuring BFD for IS-IS

Configuring BFD for IS-IS helps the system to rapidly detect the link status and speeds up IS-IS convergence in the case of a link failure.

Prerequisites

Basic IS-IS functions are configured.

Context

To accelerate IS-IS convergence speed when the link status changes, you can configure BFD on the IS-IS link. The access device supports configuration of static and dynamic BFD for IS-IS. When BFD sessions are configured in both methods, the static BFD session takes precedence over the dynamic BFD session.

- Static BFD refers to configuring BFD session parameters manually including local and remote identifiers and delivering BFD session setup requests manually.
- Dynamic BFD refers to that routing protocols dynamically trigger the establishment of BFD sessions. When setting up new neighbor relationship, routing protocols send parameters of neighbors and detection parameters (including source and destination IP addresses) to the BFD module. BFD then sets up sessions according to the received parameters between neighbors. Dynamic BFD is more flexible than static BFD.

The procedures and methods for planning and configuring IPv6 services are similar to those for configuring IPv4 services. Only some commands are different, which are described in the following configuration procedure.

 Copyright © Huawei Technologies Co., Ltd.

Procedure

Step 1 Enable BFD globally.

1. Run the **bfd** command to enable BFD globally and enter the BFD mode.
2. Run the **quit** command to quit the BFD mode.

Step 2 Create static BFD for IS-IS.

1. Run the **interface vlanif** command to enter the VLANIF interface mode.
2. Run the **isis bfd static** command to enable static BFD function on the interface.
3. Run the **quit** command to quit the VLANIF interface mode.
4. Run the **bfd bind peer-ip** command to create a BFD session.
5. Configure the discriminators.

   The local discriminator set on a device is equal to the remote discriminator set on a remote device, and the remote discriminator set on the local device is equal to the local discriminator set on the remote device. If the discriminators on the device and the remote device do not match, the session cannot be created.

   a. Run the **discriminator local** *discr-value* command to configure a local discriminator.
   b. Run the **discriminator remote** *discr-value* command to configure a remote discriminator.
6. Run the **quit** command to quit the BFD session mode.

Step 3 Create dynamic BFD for IS-IS.

- To configure BFD for all interfaces in a specified IS-IS process, perform the following steps:

1. Run the **isis** command to enter the IS-IS mode.
2. Run the **bfd all-interfaces enable** command to enable BFD for all the interfaces in a specific IS-IS process. The default parameter values are used.
3. (Optional) Run the **bfd all-interfaces** { **min-rx-interval** *receive-interval* | **min-tx-interval** *transmit-interval* | **detect-multiplier** *multiplier-value* } * command to modify the parameters of BFD session.
4. Run the **quit** command to quit the IS-IS mode.
5. (Optional) Prevent an interface from dynamically creating a BFD session.

   After BFD for IS-IS is configured, all interfaces on which neighbor relationships are Full in the IS-IS process will create BFD sessions. To prevent specific interfaces from being enabled with BFD, disable these interfaces from dynamically creating BFD sessions.

   a. Run the **interface vlanif** command to enter the VLANIF interface mode.
   b. Run the **isis bfd block** command to prevent the interface from dynamically creating a BFD session.
   c. Run the **quit** command to quit the VLANIF interface mode.

- To configure BFD for specified IS-IS interfaces, perform the following steps:

1. Run the **interface vlanif** command to enter the VLANIF interface mode.

 Copyright © Huawei Technologies Co., Ltd.

2. Run the **isis bfd enable** command to enable BFD for an IS-IS interface.
3. (Optional) Run the **isis bfd** { **min-rx-interval** *receive-interval* | **min-tx-interval** *transmit-interval* | **detect-multiplier** *multiplier-value* } * command to modify the parameters of BFD session.
4. Run the **quit** command to quit the VLANIF interface mode.

Step 4 Query the BFD session information.

Run the **display isis bfd interface** command to query the BFD configuration information of an IS-IS interface.

Run the **display isis bfd session** command to query the BFD session information.

----End

11.9.5.6 Configuring BFD for BGP or BGP4+

BFD for BGP or BGP4+ speeds up fault detection and therefore increases the route convergence speed.

Prerequisites

Basic BGP or BGP4+ functions are configured.

Context

BGP periodically sends Keepalive packets to its peers to detect the status of its peers. The detection mechanism, however, takes more than one second. When the data transmission rate reaches the level of Gbit/s, such slow detection will cause a large amount of data to be lost. As a result, the requirement for high reliability of carrier-class networks cannot be met.

BFD for BGP or BGP4+ detects faults on links between BGP peers within 50 milliseconds. The fast detection speed ensures fast BGP route convergence and minimizes traffic loss.

NOTE

By default, a multi-hop IBGP session is established between Huawei devices that set up an IBGP peer relationship. A BFD for IGP session and a BFD for IBGP session cannot be both set up between a Huawei device and a non-Huawei device that sets up a single-hop BGP session with its peer by default. In such a situation, setting up only a BFD for IGP session or a BFD for IBGP session between the Huawei and non-Huawei devices is recommended.

Procedure

Step 1 Enable BFD globally.

1. Run the **bfd** command to enable BFD globally and enter the BFD mode.

   BFD must be enabled globally before configurations relevant to BFD are performed. By default, BFD is disabled globally.
2. Run the **quit** command to quit the BFD mode.

Step 2 Create a BFD session.

 Copyright © Huawei Technologies Co., Ltd.

1. Run the **bgp** command to enter the BGP mode.
2. (Optional) Configure BFD for BGP or BGP4+ in the VPN. To configure BFD for BGP or BGP4+ for the public network, skip this step.
   - To configure BFD for BGP, run the **ipv4-family vpn-instance** *vpn-instance-name* command to enter the BGP-VPN instance IPv4 address family mode.
   - To configure BFD for BGP4+, run the **ipv6-family vpn-instance** *vpn-instance-name* command to enter the BGP-VPN instance IPv6 address family mode.
3. Run the **peer bfd enable** command to enable BFD for the peer or peer group, and a BFD session is established.

   A BFD session is set up only when the BGP session is in the Established state.

   After BFD is enabled for a peer group, BFD sessions will be created on the peers that belong to this peer group and are not configured with the **peer bfd block** command.

Step 3 (Optional) Run the **peer bfd** command to configure the BFD parameters.

The BFD parameters of peers take precedence over those of peer groups. If BFD parameters are configured on peers, they will be used in BFD session establishment.

The default intervals for transmitting BFD packets and the default detection multiplier are recommended. When changing the default values, pay attention to the network status and the network reliability requirement. A short interval for transmitting BFD packets can be configured for a link that has a higher reliability requirement. A long interval for transmitting BFD packets can be configured for a link that has a lower reliability requirement. There are three formulas:

- Actual interval for the local device to send BFD packets = max {Locally configured interval for transmitting BFD packets, Remotely configured interval for receiving BFD packets}
- Actual interval for the local device to receive BFD packets = max {Remotely configured interval for transmitting BFD packets, Locally configured interval for receiving BFD packets}
- Local detection period = Actual interval for receiving BFD packets x Remotely configured BFD detection multiplier

For example:

- On the local device, the configured interval for transmitting BFD packets is 200 ms, the interval for receiving BFD packets is 300 ms, and the detection multiplier is 4.
- On the peer device, the configured interval for transmitting BFD packets is 100 ms, the interval for receiving BFD packets is 600 ms, and the detection multiplier is 5.

Then:

- On the local device, the actual interval for transmitting BFD packets is 600 ms calculated by using the formula max {200 ms, 600 ms}; the interval for receiving BFD packets is 300 ms calculated by using the formula max {100

 Copyright © Huawei Technologies Co., Ltd.

ms, 300 ms}; the detection period is 1500 ms calculated by multiplying 300 ms by 5.

- On the peer device, the actual interval for transmitting BFD packets is 300 ms calculated by using the formula max {100 ms, 300 ms}; the interval for receiving BFD packets is 600 ms calculated by using the formula max {200 ms, 600 ms}; the detection period is 2400 ms calculated by multiplying 600 ms by 4.

wtr *wtr-value* can be specified in the command to suppress frequent BFD and BGP session flapping caused by link flapping. If a BFD session over a link goes down, it does not go Up immediately after the link recovers. Instead, the BFD session waits for the WTR timer to expire before going Up. If the link fails again before the WTR timer expires, BFD does not send a link fault message to BGP, and the BGP session status is stabilized.

Step 4 (Optional) Run the **peer bfd block** command to prevent a peer from inheriting the BFD function of the peer group to which it belongs.

If a peer joins a peer group enabled with BFD, the peer inherits the BFD configuration of the group and creates a BFD session. To prevent the peer from inheriting the BFD function of the peer group, perform this step.

The **peer bfd block** command and the **peer bfd enable** command are mutually exclusive. After the **peer bfd block** command is run, the BFD session is automatically deleted.

Step 5 Query the BFD session information.

- Run the **display bgp bfd session** command to query the BFD session between BGP peers.
- Run the **display bgp ipv6 bfd session** command to query the BFD session between BGP4+ peers.

----End

11.9.5.7 Configuring BFD for RSVP

BFD for RSVP is applied to a scenario where TE FRR is used and a Layer 2 device exists on the primary LSP between a PLR and its downstream neighbors. On a network where GR is enabled on the PLR and MP, BFD for RSVP is also recommended.

Prerequisites

The RSVP-TE tunnel is configured.

Context

By default, the interval at which RSVP Hello messages are sent is 3 seconds. The interval at which a neighbor is declared down is three times the interval at which RSVP Hello messages are sent. This allows devices to detect a fault in an RSVP neighbor at seconds level. If a Layer 2 device exists on a link between RSVP neighboring nodes, the neighboring node cannot rapidly detect the fault after the link fails, resulting in a great loss of data.

 Copyright © Huawei Technologies Co., Ltd.

BFD detects faults at millisecond level in protected links or nodes. BFD for RSVP rapidly detects faults in an RSVP neighbor, allowing packets to switch to a backup LSP rapidly.

NOTE

BFD for LSP can function properly though the forward path is an LSP and the backward path is an IP link. The forward path and the backward path must be established over the same link; otherwise, if a fault occurs, BFD cannot identify the faulty path. Before deploying BFD, ensure that the forward and backward paths are over the same link so that BFD can correctly identify the faulty path.

Perform the following steps on the two RSVP neighboring nodes between which a Layer 2 device resides:

Procedure

Step 1 Enable BFD globally.

1. Run the **bfd** command to enable BFD globally and enter the BFD mode.

   BFD must be enabled globally before configurations relevant to BFD are performed. By default, BFD is disabled globally.
2. Run the **quit** command to quit the BFD mode.

Step 2 Create a BFD session.

- If most RSVP interfaces on a node need BFD for RSVP, enable BFD for RSVP globally.

1. Run the **mpls** command to enter the MPLS mode.
2. Run the **mpls rsvp-te bfd all-interfaces enable** command to enable BFD for RSVP globally.

   After this command is run in the MPLS mode, BFD for RSVP is enabled on all RSVP interfaces except the interfaces with BFD for RSVP that are blocked.
3. (Optional) Block BFD for RSVP on the RSVP interfaces that need not BFD for RSVP.

   Run the **interface vlanif** command to enter the VLANIF interface mode, and run the **mpls rsvp-te bfd block** command to block BFD for RSVP.

- If certain RSVP interfaces on a node need BFD for RSVP, enable BFD for RSVP on the RSVP interfaces.

1. Run the **interface vlanif** command to enter the VLANIF interface mode.
2. Run the **mpls rsvp-te bfd enable** command to enable BFD for RSVP on an RSVP interface.

Step 3 (Optional) Configure the BFD parameters.

- If most RSVP interfaces on a node use the same BFD parameters, configure global BFD parameters.

1. Run the **mpls** command to enter the MPLS mode.
2. Run the **mpls rsvp-te bfd all-interfaces** command to configure global BFD parameters.

- If certain RSVP interfaces require BFD parameters different from global BFD parameters, configure BFD parameters on the RSVP interfaces.

 Copyright © Huawei Technologies Co., Ltd.

1. Run the **interface vlanif** command to enter the VLANIF interface mode.
2. Run the **mpls rsvp-te bfd** command to configure BFD parameters on an RSVP interface.

Step 4 Run the **display mpls rsvp-te bfd session** command to query the BFD session information.

----End

11.9.5.8 Configuring BFD for LAG

This section describes how to configure the BFD for LAG feature to quickly detect the status of a link aggregation group (LAG).

Prerequisites

Basic LAG functions have been configured.

Context

The BFD session for detecting an Ethernet LAG needs to be configured at both ends of the Ethernet LAG and cannot communicate with other types of BFD sessions.

After a BFD session for detecting an Ethernet LAG is created, the system creates a BFD sub-session for each member link. The discriminators of the BFD sub-sessions are determined through auto-negotiation. The discriminators of the BFD sessions are also generated by the local system. You do not need to configure the local and remote discriminators for the BFD sessions of the Ethernet LAG. The status of a BFD session in an Ethernet LAG becomes **Down** only when all BFD sub-sessions are **Down**.

Procedure

Step 1 Enable BFD globally.

1. Run the **bfd** command to enable BFD globally and enter BFD mode.
2. Run the **quit** command to quit the BFD mode.

Step 2 Create a BFD session.

1. Run the **bfd bind link-aggregation peer-ip** command to create a BFD session.
2. Run the **quit** command to quit the BFD mode.

Step 3 View the information about the BFD session.

Run the **display bfd session link-aggregation** command to view the BFD session information in the Ethernet LAG.

----End

11.9.6 BFD Reference Standards and Protocols

The following table lists the references.

 Copyright © Huawei Technologies Co., Ltd.

Document No.	Document Name	Protocol Compliance
RFC 5880	Bidirectional Forwarding Detection	• Compliant only with the synchronous mode • Compliant except for authentication According to RFC5880, the authentication field in a BFD control packet is optional. The authentication implementation of the device is in compliance with the RFC recommendation. The identifier of a BFD session is Your Discriminator carried in the packet. In a scenario where no authentication mechanism is available, if attackers detect that the IP address of the transmitter of a BFD session is the IP address of the local device and Your Discriminator of the BFD session is the same as that of the BFD session allocated by the local device, attackers can launch attacks, which will lead to abnormalities in the BFD session status and even in applications that detect the BFD session status. For a "BFD for IP" single-hop session, the TTL field carried in the IP packet is used to prevent remote attacks. If the TTL value equals 255, this packet is discarded. This method is not applicable to a "BFD for IP" multi-hop session.
RFC 5881	BFD for IPv4 and IPv6 (Single Hop)	Fully compliant
RFC 5882	Generic Application of BFD	• Compliant except for OSPF virtual connection
RFC 5883	BFD for Multihop Paths	• Compliant except for discriminator learning in outband mode • Compliant except for bidirectional connection
RFC 5884	BFD for mpls	Compliant except for the security mechanism
RFC 5885	BFD for PW	Compliant only with IP/UDP encapsulation

11.10 TWAMP

The Two-Way Active Measurement Protocol (TWAMP) is a technology that measures the round-trip performance of an IP network.

 Copyright © Huawei Technologies Co., Ltd.

11.10.1 TWAMP Introduction

This topic describes the definition, usage scenario, and benefits of Two-Way Active Measurement Protocol (TWAMP).

Definition

TWAMP is a technology that measures the round-trip performance of an IP network. It defines a method for measuring round-trip IP network performance between two TWAMP-capable devices. TWAMP measures the network delay, jitter, and packet loss rate (PLR) through UDP packets. Using which, the overall IP transmission performance can be measured by cooperation between deployed network devices.

TWAMP uses the client/server mode.

- The client establishes, starts, and stops a TWAMP session, and generates and maintains statistics about IP network performance.
- The server responds to the client's request to establish, start, and stop a TWAMP session.

Usage Scenario

As networks continue to rapidly develop and applications widely apply, more and more new services are being deployed, posing increasingly high requirements on network performance. A tool that rapidly provides statistics about IP network performance is vital.

Traditionally, NEs themselves generate and maintain statistics about IP network performance. To display statistics about entire network performance, a network management system (NMS) needs to manage multiple NEs and collect their statistics. However, in scenarios where no NMS is deployed or the NMS is incapable of collecting statistics, carriers cannot rapidly obtain network performance statistics.

To solve this problem, TWAMP is introduced. NEs themselves no longer need to generate or maintain statistics. The performance management system can easily obtain statistics about entire network performance by managing TWAMP clients only.

Benefits

Compared with the traditional tools used to collect statistics about IP network performance, TWAMP has the following advantages:

- Has low requirements on NMS capabilities. Compared with network quality analysis (NQA), TWAMP has a unified detection model and unified packet format. Its configuration method is simple and no test case needs to be designed.
- Does not require clock synchronization. Compared with IP flow performance measurement (IP FPM), TWAMP is easy to obtain and deploy.

Therefore, TWAMP is applicable to scenarios where statistics about IP network performance must be rapidly obtained but do not need to be highly accurate.

 Copyright © Huawei Technologies Co., Ltd.

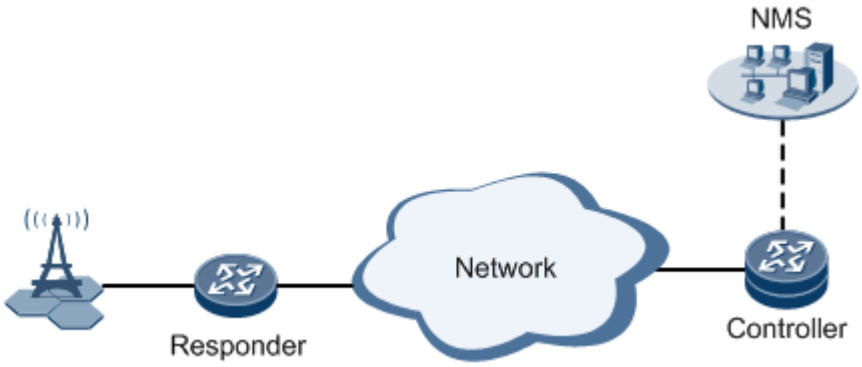
11.10.2 TWAMP Basic Concepts

This topic describes the basic concepts of the device roles and communication models of Two-Way Active Measurement Protocol (TWAMP).

Device Roles

TWAMP involves two physical devices: a Controller and a Responder. As shown in **Figure 11-46**, a device or tester works as the Controller (initiator), and the OLT or the ONT connected to an OLT works as the Responder (reflector).

Figure 11-46 Networking of TWAMP measurement



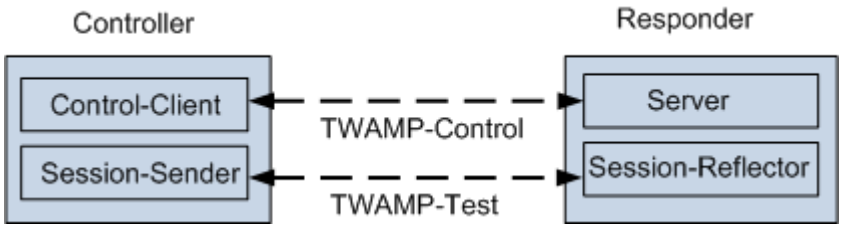
The Controller and Responder function as follows:

- Controller: Client of a TWAMP test. Serving as the initiator, it sends and receives packets over a test session, collects and calculates performance statistics, and reports these statistics to the NMS.
- Responder: Server of a TWAMP test. Serving as the receiver, it responds to packets received over a test session. It does not calculate or report performance statistics.

Communication Model

TWAMP uses client/server mode. In TWAMP Standard architecture, mapping relationships between physical devices and logical entities are shown in **Figure 11-47**. Specifically, Controller contains Control-Client and Session-Sender; Responder contains Server and Session-Reflector. The Control-Client and Server establish and start a test session. Once a test session starts, the Control-Client and Server notify the Session-Sender and Session-Reflector respectively of the session information and allow the Session-Sender to send probes and the Session-Reflector to respond to the probes.

Figure 11-47 TWAMP Standard Architecture



 Copyright © Huawei Technologies Co., Ltd.

Definition of logical roles:

- Control-Client: Establishes, starts, and stops a TWAMP session and collects statistics.
- Session-Sender: Sends probes for performance measurement after being notified by the Control-Client.
- Server: Responds to the Control-Client's request for establishing, starting, or stopping a TWAMP session.
- Session-Reflector: Responds to the probes sent by the Session-Sender after being notified by the Server.

Relationship of logical roles:

- Control-Client and Server exchange control packets through a TCP connection.
- Session-Sender and Session-Reflector exchange test packets through a UDP connection.

Protocol Packets

The Controller calculates the performance information according to fields carried in test packets. The TWAMP test packet is of the following format:

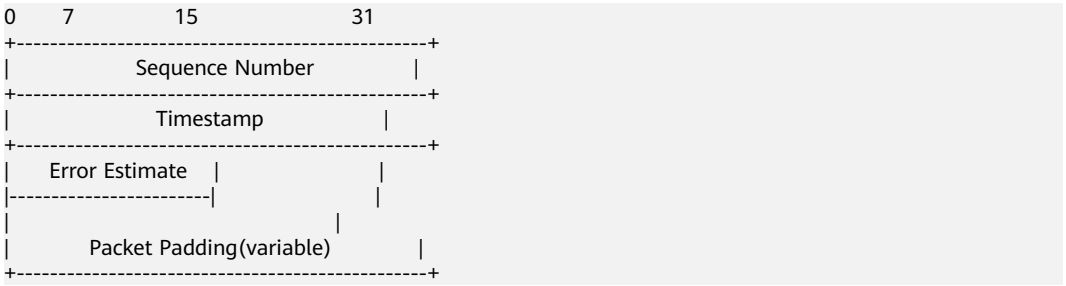


Table 11-11 Parameters of a TWAMP Test Packet

Field	Length	Description
Sequence Number	4 bytes	Sequence number (SN) of packets that is generated based on sequence of packets transmitted. It is numbered from 0 and increases with packet count. Every packet has an SN. SNs generated by the Session-reflector are irrelevant to those of packets received.
Timestamp	4 bytes	Transmission timestamp tagged on tested packets of the Session-Reflector.
Error Estimate	2 bytes	Error checks of the Session-Reflector.
Packet Padding	variable	Field for padding a packet. This field cannot be encrypted. Data fields of TWAMP test packets have the minimum 41 bytes in unauthentication mode and 104 bytes in authentication mode or encryption mode.

 Copyright © Huawei Technologies Co., Ltd.

Performance Measurement Data

TWAMP test data includes:

- Latency: Time from the point when the first bit of the packet enters the router to the point when the last bit of the packet outputs from the router.
- Jitter: Latency change. Due to network congestion, incorrect queue, or incorrect configuration, the interval at which the data packets (of even packets carried on consecutive data streams) are received on the receive is different. The change of the interval is called jitter.
- Packet loss rate (PLR): Ratio of the number of lost packets to the number of Tx packets.

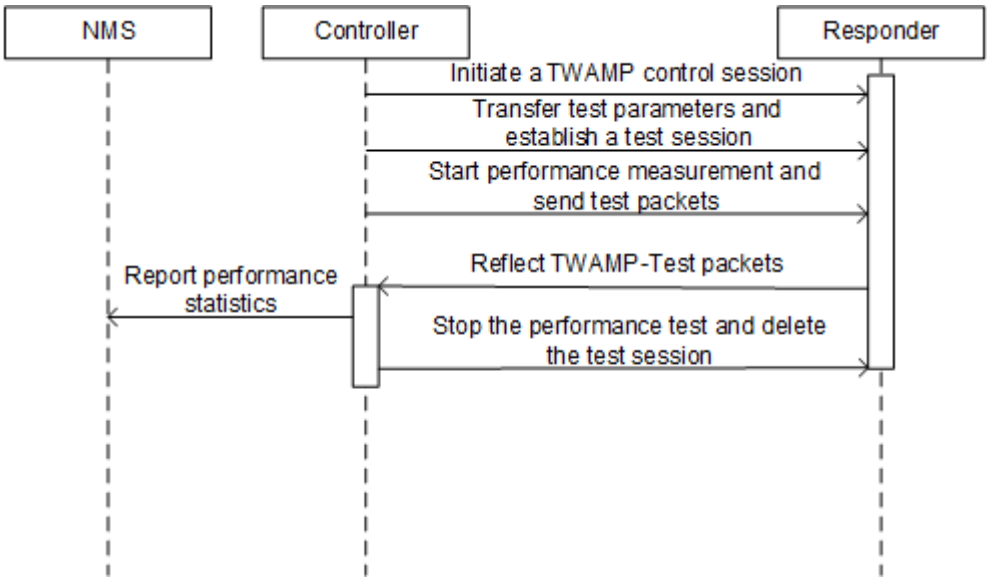
11.10.3 TWAMP Principles

This topic describes the principles and implementation process of Two-Way Active Measurement Protocol (TWAMP).

A TWAMP test covers how a control session is established, and how a test session is established, started, and stopped.

The Controller and Responder interact with each in the following way:

Figure 11-48 TWAMP test interaction



1. Control session establishment: The maintenance personnel create a control session on the configuration page of the Controller. Then, the Controller initiates a TWAMP control session.
2. Test session establishment: The Controller transmits test parameters to the Responder and the Responder initiates a test session.
3. Test session startup: The Controller starts a performance measurement test. In this test, the Controller sends the TWAMP-Test packet to the Responder and the Responder reflects the TWAMP-Test packet.
4. After receiving the reflected TWAMP-Test packet, the Controller calculates the round-trip performance information, including the packet loss rate (PLR),

 Copyright © Huawei Technologies Co., Ltd.

delay, and jitter according to the information carried in the packets, such as SN and timestamp. Then, the Controller reports the calculated performance information to the NMS.

5. Test session stop: The Controller stops the performance measurement test and deletes the Responder test session through the TWAMP control session.

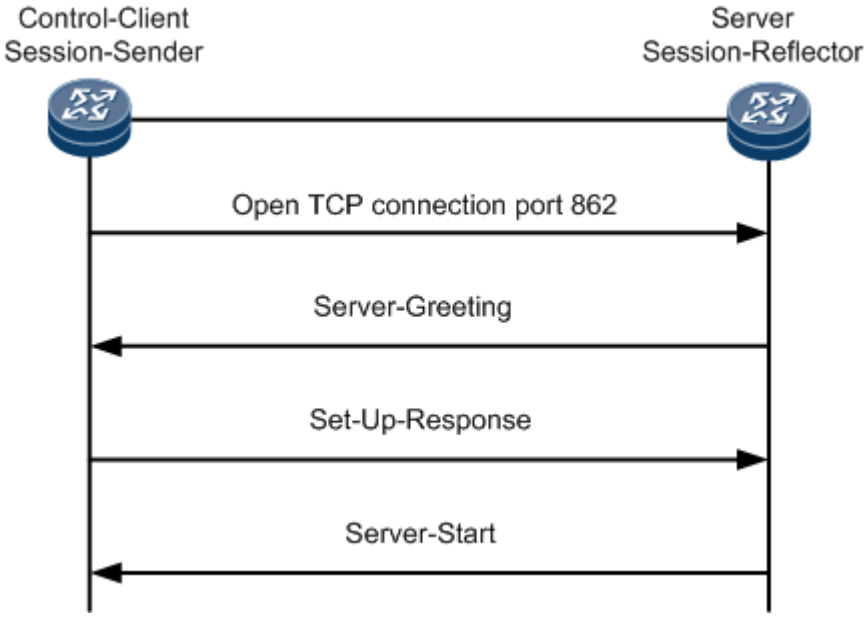
Establishing a Control Session

A control session is established based on a TCP connection for checking control layer validity and verifying parameters. A control session provides a basis for the establishment of a test session. **Figure 11-49** shows how a control session is established.

1. The Server specifies a TCP port ID (the default port ID is 862), and the Control-Client initiates a TCP connection.
2. The Server replies with a Server-Greeting message to notify the Control-Client of the server configurations.
3. After receiving the Server-Greeting message, the Control-Client sends a Set-Up-Response message to the Server to establish a control session.
4. The Server verifies the Set-Up-Response message and replies with a Server-Start message.

A control session is established between the Control-Client and the Server.

Figure 11-49 Establishing a control session



Establishing a Test Session

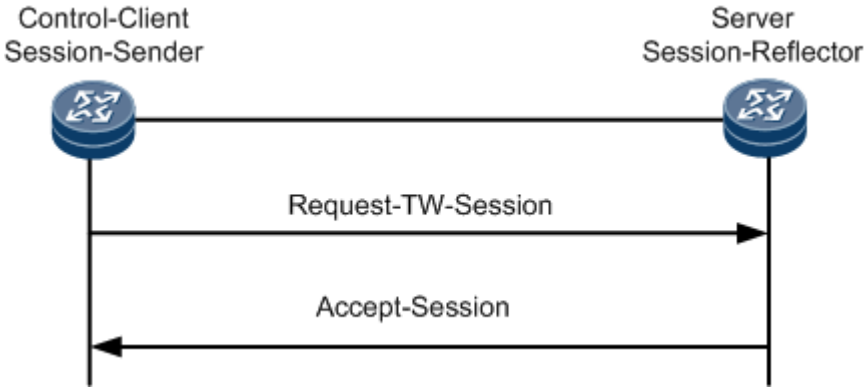
After a control session is established, users can specify an IP address and a UDP port ID on the Control-Client to establish a test session. **Figure 11-50** shows how a test session is established.

 Copyright © Huawei Technologies Co., Ltd.

1. The Control-Client sends a Request-TW-Session message carrying an IP address and a UDP port ID to the Server through the TCP connection.
2. After receiving the Request-TW-Session message, the Server establishes a test session based on the IP address and UDP port ID in the Request-TW-Session message and replies to the Control-Client with an Accept-Session message.

A test session is established between the Control-Client and the Server.

Figure 11-50 Establishing a test session



Starting a Test Session

A test session is started based on a control session. After a test session is established, the Control-Client can send a Start-Session message to the Server. After receiving the Start-Session message, the Server instructs the control session to start all test sessions that are established based on the control session. **Figure 11-51** shows how a test session is started.

1. The Control-Client sends a Start-Session message to the Server.
2. After receiving the Start-Session message, the Server notifies the Session-Reflector of the test session information, so the Session-Reflector can respond to probes.
3. The Session-Reflector replies to the Control-Client with a Start-ACK message to start the test session.
4. After receiving the Start-ACK message, the Control-Client notifies the Session-Sender of the test session information, so the Session-Sender can send probes.

The test session is started, and the Session-Reflector starts to respond to probes.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-51 Starting a test session

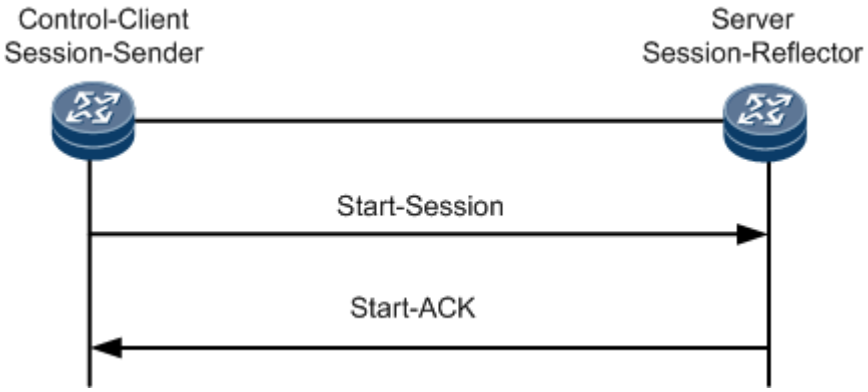
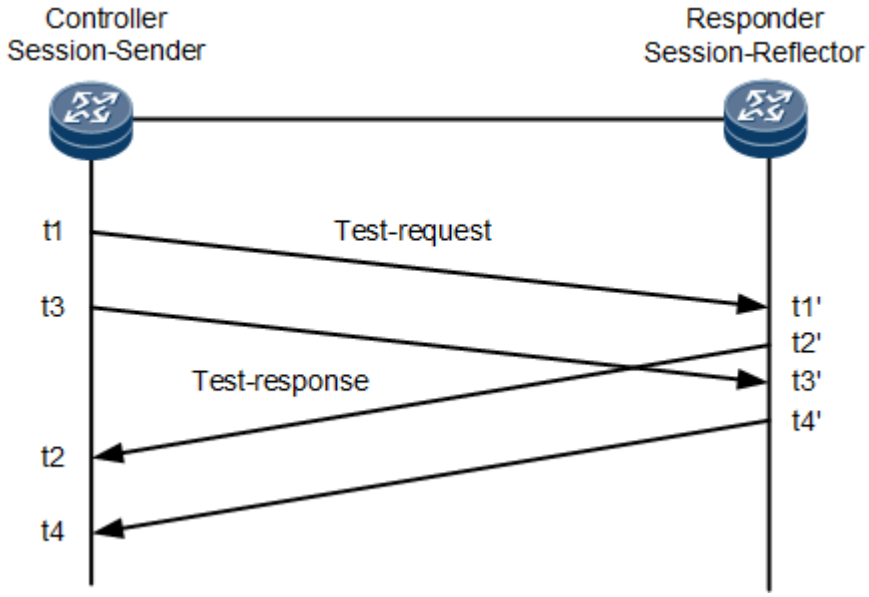


Figure 11-52 TWAMP Light performance measurement



As shown in **Figure 11-52**, after the statistical task is established, the TWAMP-Test protocol functions as the probe for sending and receiving performance measurement, and uses the preset IP address and UDP port ID of the statistical session. The Controller sends TWAMP-Test packets. After receiving the TWAMP-Test packets, the Responder reflects the packets to the Controller. The Controller collects TWAMP measurement statistics. The detailed performance measurement process is as follows:

1. After receiving the TWAMP-Test packet, the Controller calculates the round-trip packet loss rate, delay, and jitter based on the SN and timestamp in the packet.

   Delay

   The delay is calculated based on timestamps. The Controller sends a probe carrying a Tx timestamp t1, and the Responder replies with a response probe carrying an Rx timestamp t1' and a response timestamp t2'. After receiving

 Copyright © Huawei Technologies Co., Ltd.

the response probe, the Controller records the Rx timestamp t2. The delay during a single period is calculated based on 4 timestamps.

Delay1 = t2 - t1- (t2' - t1')

Jitter

The value of this parameter varies depending on the absolute value of the delay data in the adjacent period.

According to the previous delay formula, the delay value of the adjacent period is Delay2 = t4 - t3 - (t4'- t3').

Jitter = | Delay2 — Delay1 |

PLR

The controller sends a probe packet carrying the Tx SN. The Responder does not generate an independent SN. Instead, the Responder copies the Controller SN as the response SN. The final PLR is calculated based on the Rx and Tx packet data.

PLR = |Number of Tx packets - Number of Rx packets|/Number of Tx packets

2. The Controller reports the calculated performance statistics to the NMS. Then, customers can read related statistics on the NMS.

   Statistics calculated in different ways are reported to the NMS in different ways:

   - Continuous measurement: The test result is reported to the NMS through the performance monitoring (PM).
   - On-demand measurement: The test result is reported to the NMS through the MIB.

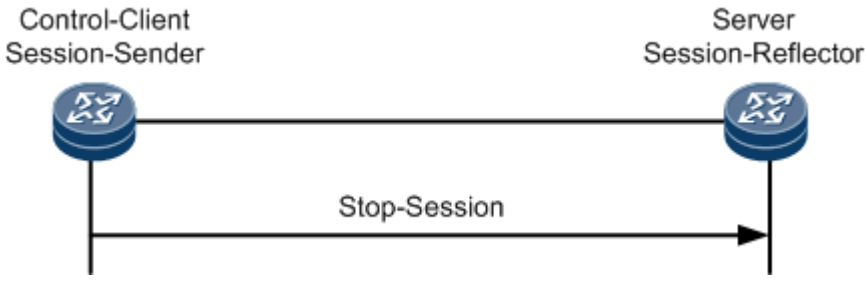
Stopping a Test Session

After statistics are collected, users can stop a test session. **Figure 11-53** shows how a test session is stopped.

1. The Control-Client sends a Stop-Session message to instruct the Server to stop collecting statistics.
2. After receiving the Stop-Session message, the Server disables the Session-Reflector from responding to probes in a test session.

The test session is stopped.

Figure 11-53 Stopping a test session



 Copyright © Huawei Technologies Co., Ltd.

11.10.4 TWAMP for ONT

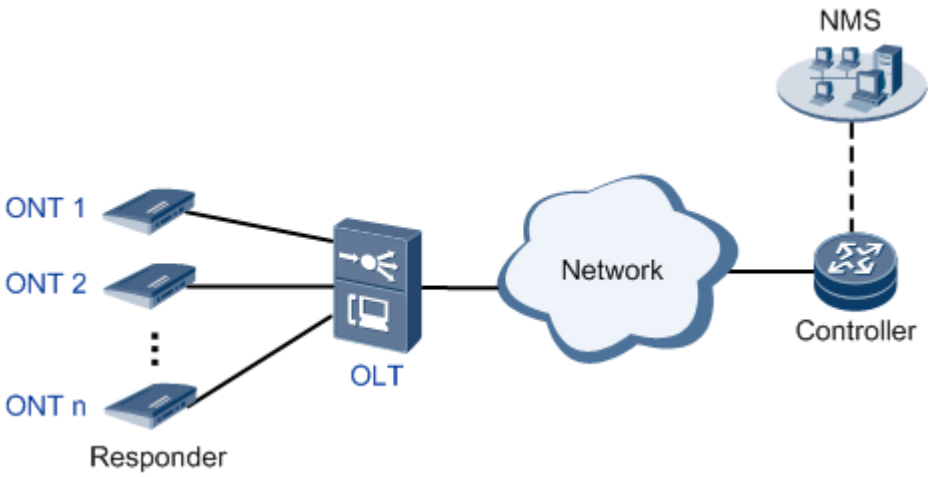
When an ONT connected to an OLT works as a responder, this ONT needs to be managed through the OLT.

Application Scenario

As shown in **Figure 11-54**, the OLT and ONT complete the Two-Way Active Measurement Protocol (TWAMP) test process together. The ONT connected to OLT works as a responder. The OLT implements the configuration functions. Through the configuration of OLT, frequent operations on a large number of ONTs can be avoided.

The TWAMP test service can be quickly restored in xPON line fault scenarios such as intermittent line disconnection, ONT reset, type B single-homing switching, and type B dual-homing switching.

Figure 11-54 TWAMP for ONT network



Logic Model

The mapping relationships between the physical devices and logical entities are shown in **Figure 11-55**

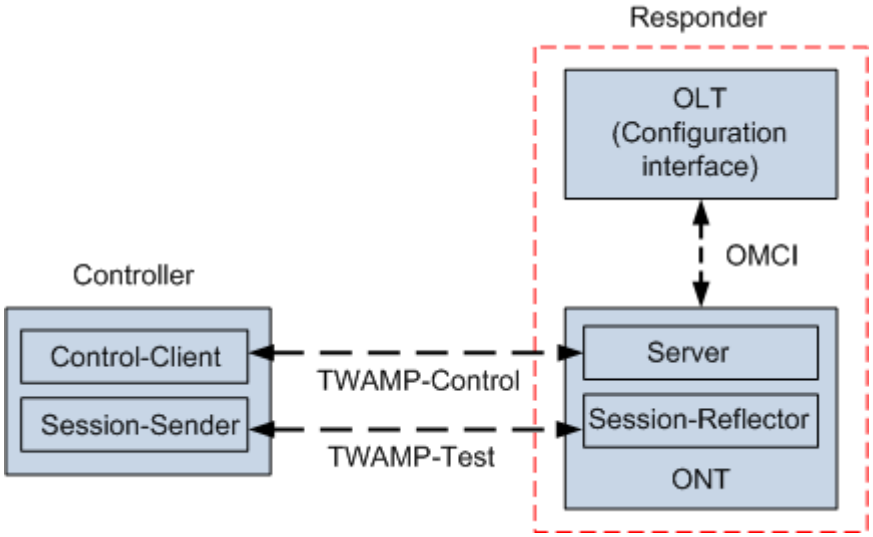
The OLT implements configuration and issues the configured parameters and running status to the ONT through OMCI.

The ONT implements control protocol and test protocol, that is, the ONT processes TWAMP control packets and test packets and interacts with the Controller logically.

Data issued by the OLT to the ONT is static data, and dynamic data is automatically generated and maintained through TWAMP control protocol interaction by the ONT.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-55 TWAMP for ONT logical model



11.10.5 Configuring TWAMP

In the TWAMP standard mode, the test parameters are flexibly negotiated by the control message between the responder and the controller. The test task is dynamically initiated.

11.10.5.1 Configuring TWAMP Responder

The control session parameters for the server role and the test session parameters for the reflector role are configured on the responder.

Prerequisites

Ensure that the Controller and Responder are routable and IP links between them work properly.

The controller creates the statistic session, initiates the TWAMP control session, and starts the test session of responder.

Procedure

Step 1 Run the **sysman server source twamp** command to bind a specified source port to the TWAMP service protocol.

NOTE

In versions earlier than V100R020C00, this step is not required. In V100R020C00 and later versions, this step must be performed to make the TWAMP function take effect.

Step 2 Run the **nqa twamp** command to create a TWAMP service, and enter TWAMP mode.

Step 3 Configure the control session parameters of TWAMP server.

1. Run the **server** command to enable the server function and enter TWAMP server mode.

 Copyright © Huawei Technologies Co., Ltd.

2. (Optional) To ensure compatibility of non-standard client devices, run the **tcp port** *port-number* [**all** | **vpn-instance** *vpn-instance-name*] command to configure a TCP dectecting port ID on the TWAMP server. The default TCP detecting port ID is 862.
3. (Optional) Run the **control-session inactive** *inactive-time* command to configure the inactive timeout duration of a TWAMP control session on the TWAMP server. If a TWAMP control session is inactive within the specified inactive timeout duration, the control session is deleted to avoid occupying system resources.
4. Run the **quit** command to return TWAMP mode.

Step 4 Configure the test session parameters of TWAMP reflector.

1. Run the **reflector** command to enable the TWAMP reflector function and enter TWAMP reflector mode.
2. (Optional) Run the **test-session inactive** *inactive-time* command to configure the inactive timeout duration of TWAMP test sessions on the TWAMP reflector. If a TWAMP test session is inactive within the specified inactive timeout duration, the test session is stopped.
3. Run the **quit** command to return TWAMP mode.

Step 5 Query the real-time statistics about TWAMP control session and TWAMP test session on a responder.

- Run the **display twamp control-session** command to query the information about control sessions on the TWAMP server.
- Run the **display twamp test-session** command to query the information about test sessions on the TWAMP reflector.
- Run the **display twamp global-info** command to query the global information about TWAMP server and reflector.

----End

Example

VLAN 10 and VLANIF 10 have been created. To set the TCP detecting port ID to 2000, the inactive timeout duration of a TWAMP control session to 600s, and the inactive timeout duration of a TWAMP test session to 600s, do as follows:

```
huawei(config)#sysman server source twamp vlanif 10    //V100R020C00 and later versions
huawei(config)#nqa twamp
huawei(config-twamp)#server
huawei(config-twamp-srv)#tcp port 2000
huawei(config-twamp-srv)#control-session inactive 600
huawei(config-twamp-srv)#quit
huawei(config-twamp)#reflector
huawei(config-twamp-ref)#test-session inactive 600
huawei(config-twamp-ref)#quit
```

11.10.5.2 Configuring TWAMP for ONT

Configuring the parameters through the MA5800 makes the ONT reflect the TWAMP test packet as the responder side.

 Copyright © Huawei Technologies Co., Ltd.

Prerequisites

Ensure that the Controller and Responder are routable and IP links between them work properly.

The controller creates the statistic session, initiates the TWAMP control session, and starts the test session of responder.

Procedure

Step 1 Run the **sysman server source** command to bind a specified source port to the TWAMP service protocol.

NOTE

In versions earlier than V100R020C00, this step is not required. In V100R020C00 and later versions, this step must be performed to make the TWAMP function take effect.

Step 2 Run the **nqa twamp ont** command to enable the TWAMP function of an ONT and enter TWAMP ONT mode.

Step 3 Run the **ont** [**eth** *ont-portid*] { **untag** | **vlan** *vlanid* [*priority1*] | **outer-vlan** *outer-vlanid* [*priority2*] **inner-vlan** *inner-vlanid* [*priority3*] } **ipaddress** *ipaddr* { *mask-length* | *mask-ipaddr* } **gateway** *gateway* command to configure the TWAMP parameters of an ONT.

Before performing a TWAMP test on an ONT, run the **ont** command to configure the TWAMP parameters of the ONT. Otherwise, the test cannot be performed.

Step 4 Configure the control session parameters of TWAMP Server on an ONT

1. Run the **server** command to enable the TWAMP server function on an ONT and enter TWAMP ONT server mode.
2. Run the **client permit source** command to configure the range of IP addresses of clients that are allowed to access the TWAMP server of an ONT. By default, the range of IP addresses of clients that are allowed to access the TWAMP server is not configured. In this case, clients with any IP addresses are not allowed to access the TWAMP server.
3. (Optional) To ensure compatibility of non-standard client devices, run the **tcp port** *port-number* [**all** | **vpn-instance** *vpn-instance-name*] command to configure a TCP detecting port ID on the TWAMP server on an ONT. The default TCP detecting port ID is 862.
4. (Optional) Run the **control-session inactive** *inactive-time* command to configure the inactive timeout duration of a TWAMP control session on the TWAMP server on an ONT. If a TWAMP control session is inactive within the specified inactive timeout duration, the control session is deleted to avoid occupying system resources.
5. Run the **quit** command to return TWAMP ONT mode.

Step 5 Configure the test session parameters of TWAMP reflector on an ONT.

1. Run the **reflector** command to enable the TWAMP reflector function on an ONT and enter TWAMP ONT reflector mode.
2. (Optional) Run the **test-session inactive** *inactive-time* command to configure the inactive timeout duration of TWAMP test sessions on the TWAMP reflector on an ONT. If a TWAMP test session is inactive within the specified inactive timeout duration, the test session is stopped.

 Copyright © Huawei Technologies Co., Ltd.

3. Run the **quit** command to return TWAMP ONT mode.

Step 6 Query the real-time statistics about TWAMP control session and TWAMP test session on an ONT responder.

- Run the **display twamp ont control-session** command to query the information about control sessions on the TWAMP server on an ONT.
- Run the **display twamp ont test-session** command to query the information about test sessions on the TWAMP reflector on an ONT.
- Run the **display twamp ont global-info** command to query the global information about TWAMP server and reflector on an ONT.

----End

Example

VLAN 10 and VLANIF 10 have been created. Enable the TWAMP function for ONT 1 connected to port 0/3/0. Configure the ONT to use the NNI port for the test. The packets are encapsulated with one VLAN tag, the VLAN ID is 10, the IP address is 192.168.1.2, the subnet mask length is 24 bits, and the gateway address is 192.168.1.1. The IP address of the client that is allowed to access is 10.10.10.2.

- On the TWAMP Responder of the ONT, set the TCP detecting port ID to 2000, the inactive timeout duration of a TWAMP control session to 600s.
- On the TWAMP Responder of the ONT, set the inactive timeout duration of the TWAMP test session to 600s.

```
huawei(config)#sysman server source twamp vlanif 10    //V100R020C00 and later versions
huawei(config)#nqa twamp ont 0/3/0 1
huawei(config-twamp-ont-0/3/0/1)#ont vlan 10 ipaddress 192.168.1.2 24 gateway 192.168.1.1
huawei(config-twamp-ont-0/3/0/1)#server
huawei(config-twamp-ont-0/3/0/1-srv)#tcp port 2000
huawei(config-twamp-ont-0/3/0/1-srv)#control-session inactive 600
huawei(config-twamp-ont-0/3/0/1-srv)#quit
huawei(config-twamp-ont-0/3/0/1)#reflector
huawei(config-twamp-ont-0/3/0/1-ref)#test-session inactive 600
```

11.10.6 TWAMP Reference Standards and Protocols

The following table lists the references.

Document	Document Name	Protocol Compliance
RFC 5357	Two-Way Active Measurement Protocol (TWAMP)	Partially compliant: • Supports the TWAMP Responder only. • Supports unauthenticated TWAMP test packets only. Does not support authentication and encryption TWAMP test packets.

11.11 TWAMP Light

Two-Way Active Measurement Protocol (TWAMP) Light is a light version of TWAMP, which is defined in RFC 5357. TWAMP Light measures the round-trip

 Copyright © Huawei Technologies Co., Ltd.

performance of an IP network by using simplified control protocols to establish test sessions.

11.11.1 TWAMP Light Introduction

This topic describes the definition and usage scenario of Two-Way Active Measurement Protocol (TWAMP) Light.

Definition

TWAMP Light is a light version of TWAMP, which is defined in RFC 5357. TWAMP Light measures the round-trip performance of an IP network by using simplified control protocols to establish test sessions.

Purpose

On conventional IP networks, a universal tool that can rapidly provide statistics about IP network performance is vital for carriers to perform operation, administration, and maintenance (OAM). Currently, Network Quality Analysis (NQA) and IP Flow Performance Measurement (IP FPM) are the tools most often used for this purpose. However, NQA requires complex deployment and does not support intercommunication between Huawei and non-Huawei devices, and IP FPM has high requirements on network devices and is applicable only to a few scenarios. To resolve this problem, the Internet Engineering Task Force IP performance monitoring (IETF IPPM) group has defined a set of protocols, including TWAMP.

TWAMP measures the round-trip performance of an IP network. It includes two versions: TWAMP Standard and TWAMP Light.

- TWAMP Standard: Involves control packet exchange and test packet exchange. In this version, two devices can flexibly negotiate test parameters through control packets and dynamically initiate test tasks. However, protocol implementation is complex.
- TWAMP Light: Involves only test packet exchange (TWAMP Light does not involve control packet exchange). In this version, test parameters are manually configured on two devices and test tasks are manually issued. However, protocol implementation is simple. TWAMP Light simplifies the working process of protocols and is easier to deploy in actual application.

11.11.2 TWAMP Light Basic Concepts

This topic describes the basic concepts of device roles and communication models of TWAMP, and the differences between TWAMP Standard and TWAMP Light.

Device Roles

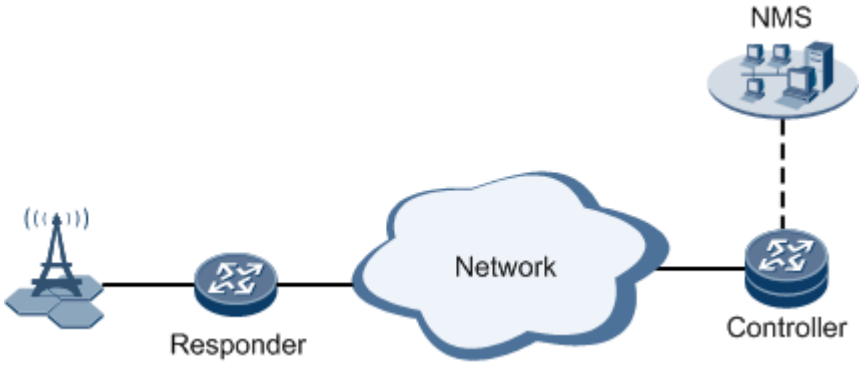
TWAMP involves two physical devices: a Controller and a Responder. As shown in **Figure 11-56**, a device or tester works as the Controller (initiator), and the MA5800 works as the Responder (sender).

 Copyright © Huawei Technologies Co., Ltd.

NOTE

The MA5800 works as the Responder only in TWAMP Light.

Figure 11-56 Networking of TWAMP measurement



The Controller and Responder function as follows:

- Controller: Serves as the initiator of a TWAMP test session, which sends and receives packets over a test session, collects and calculates performance statistics, and reports these statistics to the NMS.
- Responder: Serves as the receiver of a TWAMP test session, which responds to packets received over a test session. It does not calculate or report performance statistics.

Communication Models

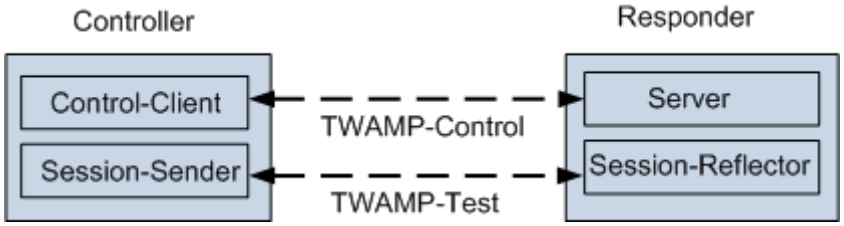
For details on the communication models and logical entities, see **Communication Model** in "TWAMP Basic Concepts."

In the TWAMP Standard and TWAMP Light architectures, physical devices and logical entities have different mapping relationships.

- TWAMP Standard

  Mapping relationships between physical devices and logical entities in TWAMP Standard architecture are shown in **Figure 11-57**. Specifically, Controller contains Control-Client and Session-Sender; Responder contains Server and Session-Reflector.

  Figure 11-57 TWAMP Standard architecture

- TWAMP Light

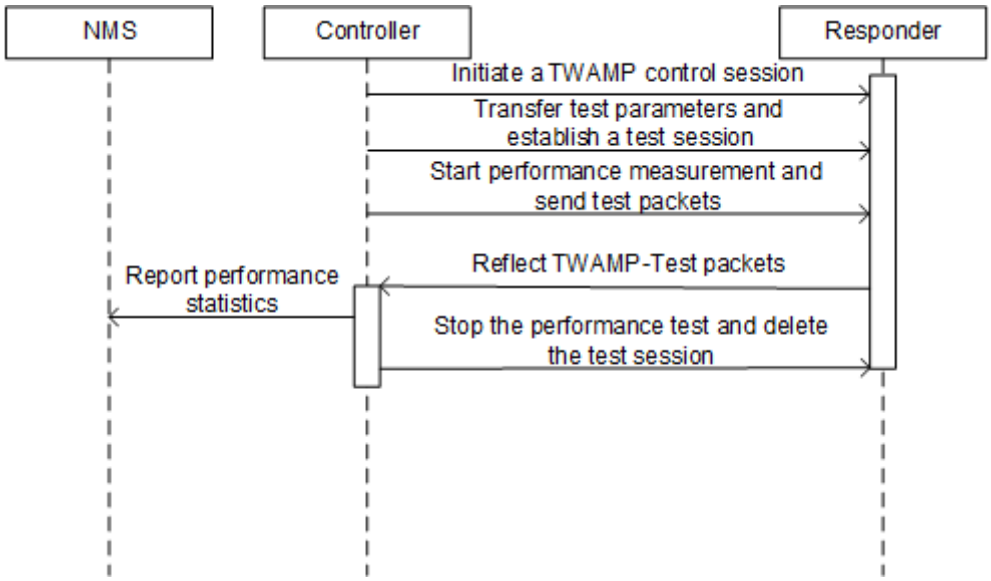
  Mapping relationships between physical devices and logical entities in TWAMP Light architecture are shown in **Figure 11-58**. Specifically, Controller

 Copyright © Huawei Technologies Co., Ltd.

contains Control-Client, Session-Sender, and Server; Responder contains only Session-Reflector.

Unlike TWAMP Standard, TWAMP Light has parameters statically configured for a test session. You can configure the IP address and UDP port ID on the Responder by running commands. After a test session is created, the performance statistics, such as the packet loss rate, delay, and jitter can be calculated based on the exchange of TWAMP-Test packets. No control session (required in TWAMP Standard) needs to be set up during performance detection. TWAMP Light does not need any control protocol for parameter negotiation. It simplifies the communication model of TWAMP and greatly reduces requirements on the Responder performance, allowing the Responder to be rapidly deployed. In addition, TWAMP Light supports plug-and-play.

Figure 11-58 TWAMP Light architecture



11.11.3 TWAMP Light Principles

This topic describes the principles and implementation process of Two-Way Active Measurement Protocol (TWAMP) Light.

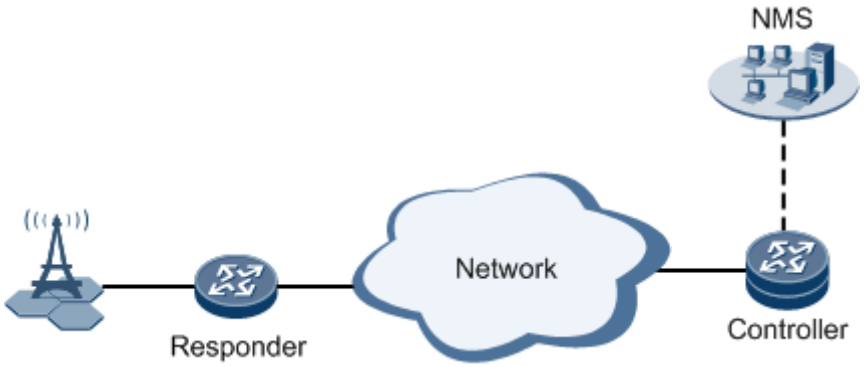
TWAMP Light Principles

A TWAMP Light service must be established before TWAMP Light is implemented.

1. **Establishing a TWAMP Light service**

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-59 Establishing a TWAMP Light service



The Controller and Responder are deployed as shown in **Figure 11-59**.

a. The Control-Client on the Controller creates a TWAMP Light test session.
b. The Responder is specified as the Session-Reflector.
c. The Session-Sender on the Controller starts the test session, and the Controller sends TWAMP-Test packets to the Responder based on the configured packet sending rate and packet profile.
d. The Responder responds to the TWAMP-Test packets.

2. **Measuring performance**

   TWAMP Light supports both on-demand measurement and proactive measurement.

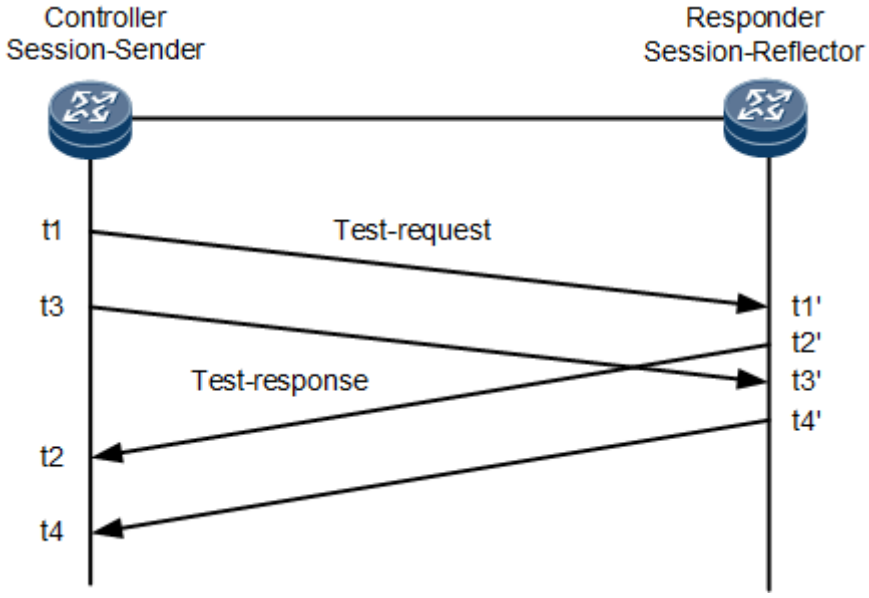
   – On-demand measurement works within a specified period after being started. It can be performed once or periodically within the specified period.
   – Proactive measurement works continuously after being started to collect statistics.

   TWAMP Light defines two types of TWAMP-Test packets: Test-request packets and Test-response packets.

   – Test-request packets are sent from the Controller to the Responder.
   – Test-response packets are sent as replies from the Responder to the Controller.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-60 TWAMP Light implementation



In **Figure 11-60**, TWAMP-Test packets function as probes and carry the IP address and UDP port ID that are predefined for the test session between the Controller and Responder. The Controller sends a TWAMP-Test packet to the Responder, and the Responder replies to it. The Controller collects TWAMP statistics as follows:

a. The Controller collects statistics about the round—trip delay, jitter, and packet loss rate based on the sequence numbers and timestamps carried in TWAMP-Test packets.

   Delay

   The delay is calculated based on timestamps. The Controller sends a probe carrying a sending timestamp t1, and the Responder replies with a response probe carrying a receiving timestamp t1' and a responding timestamp t2'. After receiving the response probe, the Controller records the receiving timestamp t2. The delay during a single period is calculated based on the four timestamps.

   Delay1 = t2 - t1- (t2' - t1')

   Jitter

   The jitter is calculated based on two consecutive delays.

   Based on the preceding delay formula, the delay of the neighboring measurement interval can be calculated as follows: Delay2 = t4 - t3 - (t4' - t3')

   Jitter = | Delay2 — Delay1 |

   Packet loss rate

   The packet loss rate is calculated based on the sequence numbers carried in probes. The Controller sends a probe with a sequence number, and the Responder replies with a response probe with the same sequence number. The packet loss rate is calculated as follows:

   Packet loss rate = |Number of sent packets - Number of received packets|/Number of sent packets

 Copyright © Huawei Technologies Co., Ltd.

b. The Controller reports the calculated performance statistics to the NMS. Then, customers can read related statistics on the NMS.

Statistics calculated in different ways are reported to the NMS in different ways:

- Continuous measurement: The test result is reported to the NMS through the performance monitoring (PM).
- On-demand measurement: The test result is reported to the NMS through the MIB.

11.11.4 Configuring TWAMP Light Responder

Users can create a test session on the Two-Way Active Measurement Protocol (TWAMP) Light Responder for responding to TWAMP-Test packets.

Prerequisites

Ensure that the controller and responder are routable and IP links between them work properly.

Procedure

Step 1 Run the **sysman server source** command to bind a specified source port to the TWAMP service protocol.

NOTE

In versions earlier than V100R020C00, this step is not required. In V100R020C00 and later versions, this step must be performed to make the TWAMP function take effect.

Step 2 Run the **nqa twamp-light** command to create a TWAMP Light service, and enter TWAMP Light mode.

Step 3 Run the **responder** command to enable the TWAMP Light responder function, and enter TWAMP Light responder mode.

Step 4 Run the **test-session** *session-id* **local-ip** *local-ip-address* **remote-ip** *remote-ip-address* **local-port** *local-port-number* **remote-port** *remote-port-number* [**vpn-instance** *vpn-instance-name*] [**description** *description-value*] command to create a session on the responder of a TWAMP Light service.

Step 5 Run the **display twamp-light responder test-session** command to query the real-time session statistics of a TWAMP Light responder.

----End

Example

VLAN 10 and VLANIF 10 have been created. Assume that the session ID is 1, IP address of the responder is 10.1.1.1, IP address of the sender is 10.2.2.2, UDP port IDs of the responder and sender are 3000, and the VPN instance name is vpn1. To create a session on a responder of the TWAMP Light service with the preceding parameters, do as follows:

```
huawei(config)#sysman server source twamp vlanif 10    //V100R020C00 and later versions
huawei(config)#nqa twamp-light
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-twamp-light)#responder
huawei(config-twamp-light-responder)#test-session 1 local-ip 10.1.1.1 remote-ip 10.2.2.2 local-port
3000 remote-port 3000 vpn-instance vpn1
huawei(config-twamp-light-responder)#quit
huawei(config-twamp-light)#quit
huawei(config)#display twamp-light responder test-session
  Total number : 1
  ------------------------------------------------------------------------
  ID     Local-IP        Local-Port     Remote-IP       Remote-Port
  ------------------------------------------------------------------------
   1     10.1.1.1           3000        10.2.2.2           3000
  ------------------------------------------------------------------------
```

11.11.5 TWAMP Light Reference Standards and Protocols

The following table lists the references.

Document	Document Name	Protocol Compliance
RFC 5357	Two-Way Active Measurement Protocol (TWAMP)	Partially compliant: • Supports the TWAMP Light Responder only. • Supports unauthenticated TWAMP test packets only. Does not support authentication and encryption TWAMP test packets.

11.12 Routing

Routing is a common term used for describing the path through which the packets from a host in a network travel to a host in another network.

11.12.1 IP Routing Overview

11.12.1.1 Introduction to Routing

Definition

Routing is a common term used for describing the path through which the packets from a host in a network travel to a host in another network.

Routers send packets on the Internet. A router selects a suitable path in a network according to the destination address included in a received packet, and sends the packet to the next router on the path. In this way, the packet travels over the Internet Until it reaches the destination host.

Purpose

The access equipment, serving as a basic element in the entire telecom network, must support the functions of remote operation, management and maintenance on the equipment itself.

 Copyright © Huawei Technologies Co., Ltd.

With the development of small-size access equipment that can be managed remotely, the access equipment needs to feature the functions of a BRAS, such as allocation of network addresses and user management. In this way, the access equipment must support the routing feature.

A MA5800 can also serve as a router.

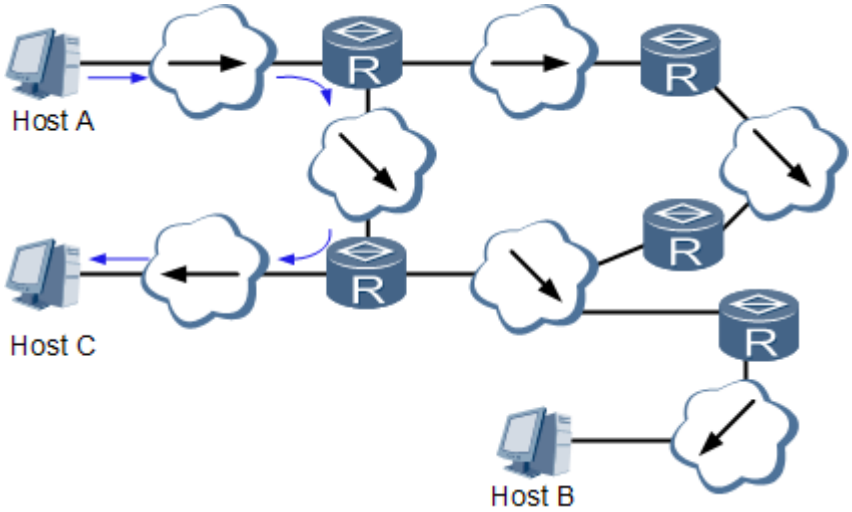
11.12.1.2 Routers

On the Internet, network connection devices control network traffic and ensure data transmission quality on networks. Common network connection devices include hubs, bridges, switches, and routers.

As a standard network connection device, a router is used to select routes and forward packets. Based on the destination address in the received packet, a router selects a path to send the packet to the next router. The last router is responsible for sending the packet to the destination host. In addition, a router can select an optimal path for data transmission.

For example, in **Figure 11-61**, traffic from Host A to Host C needs to pass through three networks and two routers. The hop count from a router to its directly connected network is zero. The hop count from a router to a network that the router can reach through another router is one. The rest can be deduced by analogy. If a router is connected to another router through a network, a network segment exists between the two routers, and they are considered adjacent on the Internet. In the figure, the arrows indicate network segments. The routers do not need to know about the physical link composition of each network segment.

Figure 11-61 Network segment and hop count



Network sizes may vary greatly, and the actual lengths of network segments vary as well. Therefore, you can set a weighted coefficient for the network segments of each network and then measure the cost of a route based on the number of network segments.

A route with the minimal network segments is not necessarily optimal. For example, a route passing through three high-speed Local Area Network (LAN) network segments may be a better choice than one passing through two low-speed Wide Area Network (WAN) network segments.

 Copyright © Huawei Technologies Co., Ltd.

11.12.1.3 Routing Table and FIB Table

Each router maintains one routing table and one FIB table at least. A router uses the routing table to select routes, and uses the FIB table to guide packet forwarding. In Equal and Weighted Cost Multi-Path (ECMP), according to different states of the network, the traffic to the same destination network can be distributed among multiple equal-cost paths to reduce the network load or to implement the link backup function.

- Routes discovered by the various routing protocols are stored in the routing table. The routes in the routing table are divided, according to their sources, into the following types:
  - Directly connected route or interface route: is the route discovered by the link layer protocols.
  - Static route: is the route manually configured by the network administrator.
  - Dynamic route: is the route discovered by dynamic routing protocols.
- Each entry in the FIB table contains the physical or logical interface through which a packet is sent to a network segment or host to reach the next router. An entry also indicates whether the packet can be sent directly to a destination host in a directly connected network.

Routing table

Each routing entry in the routing table specifies the physical port through which a packet is sent to a subnet or a host, to the next router on the path, or to the destination host on the directly connected network without passing through another router.

Each router maintains the protocol routing table for each type of protocol and a local core routing table (or routing management table).

- Protocol routing table

  A protocol routing table stores the routing information discovered by the protocol. A routing protocol can import and advertise the routes that are discovered by other protocols. For example, if a router that runs the Open Shortest Path First (OSPF) protocol needs to use OSPF to advertise direct routes, static routes, or Intermediate System-Intermediate System (IS-IS) routes, the router must import the routes into the OSPF routing table.
- Local core routing table

  A router uses the local core routing table to store protocol routes and preferred routes. The router then delivers the preferred routes to the FIB table for forwarding packets. The router selects routes according to the priorities of protocols and costs stored in the routing table. You can run the **display ip routing-table** command to view the local core routing table of a router.

  A router maintains a local core routing table for each Virtual route forwarding instance (VRF).

The key entries of the routing table are shown in **Table 11-12**.

 Copyright © Huawei Technologies Co., Ltd.

Table 11-12 Key entries of the routing table

Entry	Description
Destination	The destination address is a 32-bit character that labels the destination IP address or destination network of an IP packet.
Mask	The mask is used with the destination address to identify the subnet address of the destination host or router. The network mask is composed of several consecutive 1s. These 1s can be expressed in either the dotted decimal notation or the number of consecutive 1s in the mask. For example, the network mask can be expressed either as 255.255.255.0 or 24. The network address of the destination host or router is obtained through the "AND" operation on the destination address and network mask. For example, if the destination address is 10.1.1.1 and the mask is 255.255.255.0, the address of the network where the host or router resides is 10.1.1.0.
Proto	Indicates the protocol through which routes are learned.
Pre	Indicates the preference added to the IP routing table for a route. To the same destination, multiple routes with different next hops and outgoing interfaces exist. The routes in the table are those discovered by different routing protocols or tare the manually configured static routes. The router selects the route with the highest preference (the smallest value) as the optimal route. For more information on the preference of each protocol, see **Table 11-15** in the Route Protocols.
Cost	Indicates the route cost. When multiple routes to the same destination have the same preference, the route with the lowest cost is selected as the optimal route. **NOTE** The Preference value is used to compare the preferences of various routing protocols, while the Cost value is used to compare the preferences of different routes of the same routing protocol.
NextHop	Indicates the IP address of the next device that an IP packet passes through.
Interface	Indicates the outgoing interface through which an IP packet is forwarded.

Based on the destination, routes can be classified as:

- Subnet route: Its destination is a subnet.
- Host route: Its destination is a host.

Based on the connection between the destination and the router, routes can be classified as:

- Direct route: Its destination network is directly connected to the router.

 Copyright © Huawei Technologies Co., Ltd.

- Indirect route: Its destination network is not directly connected to the router.

To avoid large routing tables, a default route can be assigned. Once a packet fails to find a dedicated route in the routing table, the default route is selected for forwarding the packet.

In the networks shown in **Figure 11-62**, the digits in each network represent the IP address of the network. Router 8 is connected to three networks. Therefore, it has three IP addresses and three physical ports. **Figure 11-62** is its routing table.

Figure 11-62 Interconnected networks

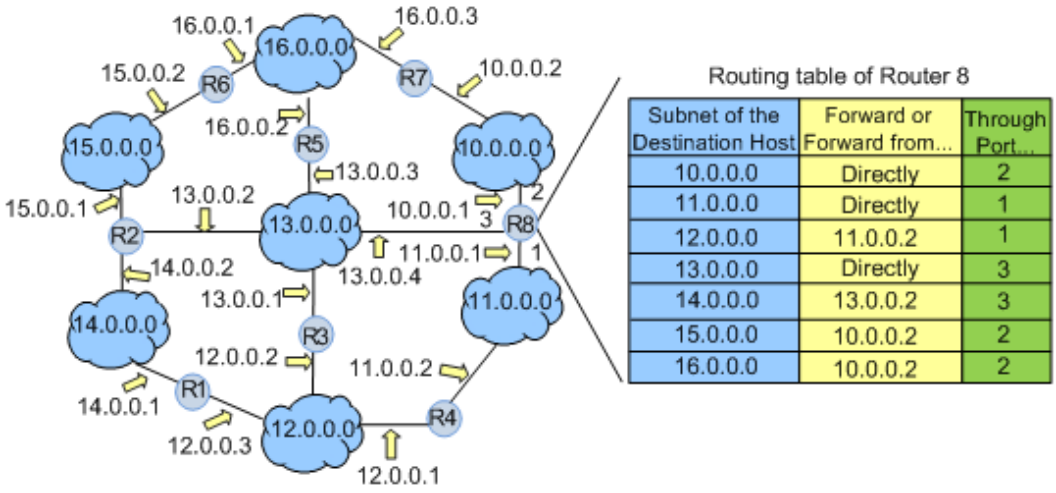


Table 11-13 shows the routing table of router 8.

Table 11-13 Routing table of Router 8

Subnet of the Destination Host	Forward or Forward from...	Through Port...
10.0.0.0	Directly	2
11.0.0.0	Directly	1
12.0.0.0	11.0.0.2	1
13.0.0.0	Directly	3
14.0.0.0	13.0.0.2	3
15.0.0.0	10.0.0.2	2
16.0.0.0	10.0.0.2	2

FIB Table

After route selection, routers send the active routes in the routing table to the FIB table. When a router receives a packet, the router searches the FIB table for the optimal route to forward the packet.

 Copyright © Huawei Technologies Co., Ltd.

The router performs the "AND" operation on the destination address in the packet and the network mask of each entry in the FIB table. The router then compares the result of the "AND" operation with the entries in the FIB table to find a match. The router chooses the optimal route to forward packets according to the best or "longest" match.

As an example, a certain router has the following brief routing table:

```
Routing Tables:
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface
        0.0.0.0/0   Static  60   0           D   120.0.0.2       Vlanif10
        8.0.0.0/8   RIP     100  3           D   120.0.0.2       Vlanif10
       10.0.0.0/8   OSPF    10   50          D   20.0.0.2        Vlanif10
      10.1.0.0/16   RIP     100  4           D   120.0.0.2       Vlanif20
      20.0.0.0/8    Direct  0    0           D   20.0.0.1        Vlanif20
```

After receiving a packet that carries the destination address 10.1.2.1, the router searches the following table:

```
 FIB Table:
 Total number of Routes : 5
Destination/Mask    Nexthop         Flag TimeStamp     Interface       TunnelID
0.0.0.0/0           120.0.0.2       SU   t[37]         Vlanif10        0x0
8.0.0.0/8           120.0.0.2       DU   t[37]         Vlanif10        0x0
10.0.0.0/8          20.0.0.2        DU   t[9992]       Vlanif10        0x0
10.1.0.0/16         120.0.0.2       DU   t[9992]       Vlanif20        0x0
20.0.0.0/8          20.0.0.1        U    t[9992]       Vlanif20        0x0
```

First, AND operation is performed on the destination address 10.1.2.1 and the masks 0, 8, and 16 to obtain the network segment addresses: 0.0.0.0/0, 10.0.0.0/8, and 10.1.0.0/16. The three addresses match the three entries in the FIB table, that is, 0.0.0.0/0 (0 bits matching), 10.0.0.0/8 (8 bits matching), and 10.1.0.0/16 (16 bits matching). According to the longest match, the VRP selects the entry 10.1.0.0/16 and forwards the packet with its destination address as 10.1.2.1 through VLANIF 20.

11.12.1.4 Routing Protocols

The MA5800 supports the configuration of static routes and the dynamic routing protocols such as RIP, OSPF, IS-IS, and BGP. The MA5800 manages the static and dynamic routes in a unified manner. The static routes and the routes discovered by the routing protocols can be shared.

- Static routes can be easily configured on a system and have lower system requirements. Static routes are applicable to simple, stable, and small-scale networks. Static routes, however, cannot automatically adapt to changes in the network topology, so they must be manually configured.
- On the other hand, dynamic routing protocols use routing algorithms to automatically adapt to changes in network topology. Dynamic routes are applicable to the network that is equipped with Layer 3 devices. The dynamic route configuration, however, has a higher requirement (such as large memory capacity) for system performance and occupies more network resources.

Classification of Dynamic Routing Protocols

Dynamic routing protocols are classified according to the following factors:

Routing protocols are classified according to the application range:

 Copyright © Huawei Technologies Co., Ltd.

- Interior Gateway Protocol (IGP): runs inside an AS, such as RIP, OSPF, and IS-IS.

  Table 11-14 Differences among the three typical IGPs

Item	RIP	OSPF	IS-IS
Protocol type	IP layer protocol	IP layer protocol	Link layer protocol
Application scope	Applies to small networks with simple architectures, such as campus networks.	Applies to medium-sized networks, such as enterprise networks.	Applies to large networks, such as Internet service provider (ISP) networks.
Routing algorithm	Uses a distance-vector algorithm and exchanges routing information over the User Datagram Protocol (UDP).	Uses the shortest path first (SPF) algorithm to generate a shortest path tree (SPT) based on the network topology, calculates shortest paths to all destinations, and exchanges routing information over IP.	Uses the SPF algorithm to generate an SPT based on the network topology, calculates shortest paths to all destinations, and exchanges routing information over IP. In IS-IS, the SPF algorithm runs separately in Level-1 and Level-2 databases.
Route convergence speed	Slow	Less than 1 second	Less than 1 second
Scalability	Not supported	Supported by partitioning a network into areas	Supported by defining levels

- Exterior Gateway Protocol (EGP): runs between different ASs, such as BGP.

Routing protocols are classified according to the type of algorithm they use:

- Distance-Vector Routing Protocol: includes RIP and BGP (BGP is also called Path-Vector).
- Link-State Routing Protocol: includes OSPF and IS-IS.

Routing Protocols and Routing Preference

Routing protocols (including the static route) can learn different routes to the same destination, but not all routes are optimal. Only one routing protocol at one time determines the optimal route to a destination. To select the optimal route,

 Copyright © Huawei Technologies Co., Ltd.

each routing protocol (including the static route) is configured with a preference (the smaller the value, the higher the preference). When multiple routing information sources coexist, the route with the highest preference is selected as the optimal route (the smaller the value is, the higher the preference is).

Table 11-15 lists various routing protocols and the default priorities of the routes discovered by them. "0" indicates the direct route, and "255" indicates any route from an untrusted source.

Table 11-15 Default priorities of routing protocols

Routing Protocol or Route Type	Default Routing Preference
DIRECT	0
OSPF	10
IS-IS	15
INTERNAL EIGRP	50
STATIC	60
RIP	100
OSPF ASE (AS-External)	150
OSPF NSSA (Not-So-Stubby Area)	150
EXTERNAL EIGRP	160
IBGP	255
EBGP	255
UNKNOWN	255

You can define the priorities for all dynamic routing protocols except the direct route (DIRECT) and the BGP (IBGP, EBGP). In addition, the priorities of any two static routes can be different.

If different routing protocols are configured with the same preference, the system determines which routes discovered by these routing protocols become the preferred routes through an internal preference. Table 11-16 shows the internal preferences of routing protocols.

Table 11-16 Internal preferences of routing protocols

Routing Protocol or Route Type	Internal Routing Preference
DIRECT	0
OSPF	10
IS-IS Level-1	15

 Copyright © Huawei Technologies Co., Ltd.

Routing Protocol or Route Type	Internal Routing Preference
IS-IS Level-2	18
STATIC	60
RIP	100
OSPF ASE (AS-External)	150
OSPF NSSA (Not-So-Stubby Area)	150
IBGP	200
EBGP	20

For example, two routes, an OSPF route and a static route, can reach the destination 10.1.1.0/24, and the preferences of both routes are set to 5. In this case, the VRP determines the optimal route according to the internal preferences listed in **Table 11-15**. The internal preference value 10 of OSPF is higher than the internal preference value 60 of the static route. Therefore, the system selects the route discovered by OSPF as the optimal route.

Route Sharing through Route Policy

Different routing protocols can find different routes as they use different algorithms. A routing protocol might need to import routes discovered by other protocols to diversify its own routes. The MA5800 supports importing the routes discovered by one protocol to another protocol. Each protocol has its own route importing mechanism.

However, a protocol only needs to import qualified routes by setting attributes of the routes to be imported. To realize a route policy, you must define the attributes of the routes to which the route policy is to be applied, such as the destination address, and the address of the router distributing routes. You can define the matching rules in advance so that they can be applied in a route policy for route distribution, reception and importing.

The two applications of the routing policy are as follows:

- When importing routes discovered by other protocols, a routing protocol can apply this filter to obtain the required routes.
- When transmitting or receiving routes, a routing protocol can apply the filter so that only the required routes are transmitted or received.

11.12.1.5 Priority-based Route Convergence

Definition

Priority-based route convergence is an important technology that improves network reliability. It provides faster route convergence for key services. For example, to minimize the interruption of key services in case of network faults, real-time multicast services require that the routes to the multicast source quickly

 Copyright © Huawei Technologies Co., Ltd.

converge, and the Multiprotocol Label Switching (MPLS) VPN bearer network requires that routes between PEs also quickly converge.

Convergence priorities provide references for the system to converge routes for service forwarding. Different routes can be set with different convergence priorities, which can be identified as critical, high, medium, and low listed in descending order.

Purpose

With the integration of network services, requirements on service differentiation increase. Carriers require that the routes for key services, such as Voice over IP (VoIP) and video conferencing services converge faster than those for common services. Therefore, routes need to converge based on their convergence priorities to improve network reliability.

Route Convergence Priority

Table 11-17 lists the default convergence priorities of public network routes. You can set convergence priorities for routes based on the requirements of a live network.

Table 11-17 Default convergence priorities of public network routes

Routing Protocol or Route Type	Convergence Priority
DIRECT	critical
STATIC	medium
32-bit host routes of OSPF and IS-IS	medium
OSPF route (except 32-bit host routes)	low
IS-IS route (except 32-bit host routes)	low
RIP	low
BGP	low

NOTE

For VPN route priorities, only 32-bit host routes of OSPF and IS-IS are identified as medium, and the other routes are identified as low.

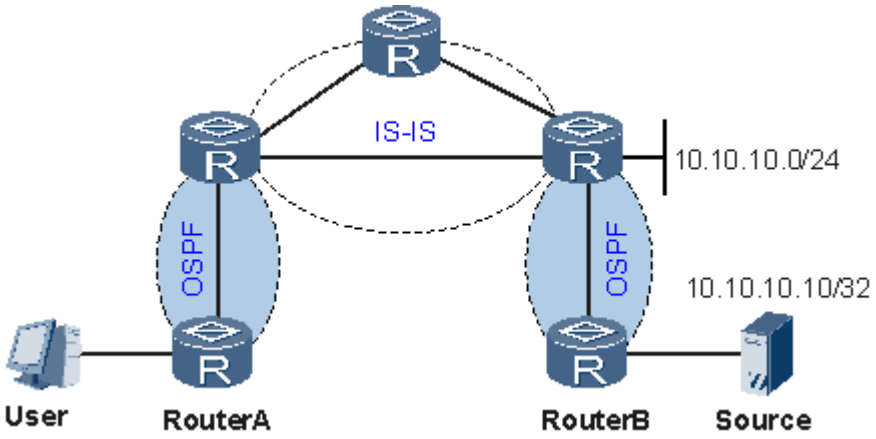
Applications

Figure 11-63 shows networking for multicast services. An IGP runs on the network; Router A is the receiver, and Router B is the multicast source server with IP address 10.10.10.10/32. The route to the multicast source server is required to

 Copyright © Huawei Technologies Co., Ltd.

converge faster than other routes, such as 10.10.10.0/24. In this case, you can set a higher convergence priority for 10.10.10.10/32 than that of 10.10.10.0/24. Then, when routes converge on the network, the route to the multicast source server 10.10.10.10/32 converges first, ensuring the transmission of multicast services.

Figure 11-63 Networking for priority-based route convergence



11.12.1.6 Load Balancing and Route Backup

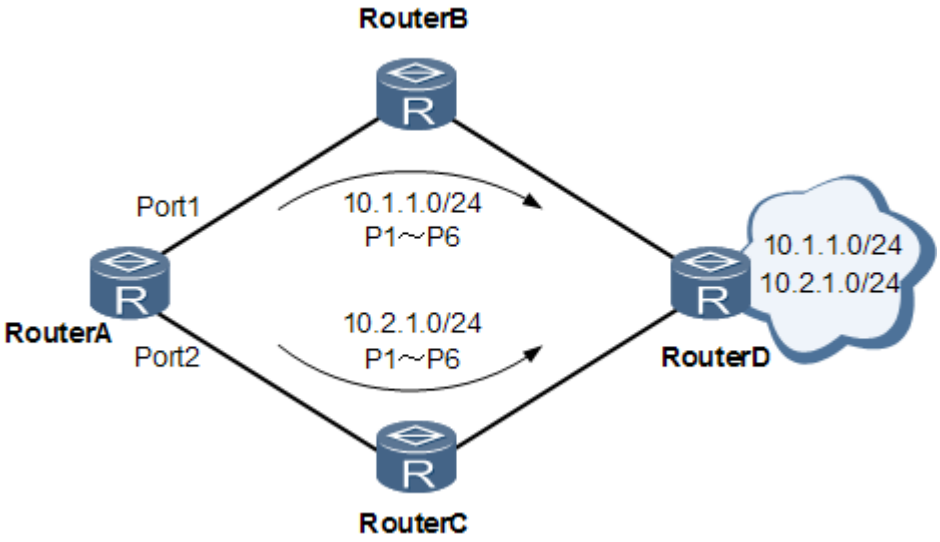
Load Balancing

The MA5800 supports the multi-route model (multiple routes with the same destination and priority). Routes discovered by one routing protocol with the same destination and cost can load-balance traffic. In each routing protocol view, you can run the **maximum load-balancing** *number* command to perform load balancing. The MA5800 adopts per-destination load balancing.

After per-destination load balancing is configured, the Router forwards packets based on the quintuple (the source address, destination address, source port, destination port, and protocol in the packets). When the quintuple is the same, the Router always chooses the next hop address that is the same as the last one to send packets. **Figure 11-64** per-destination load balancing.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-64 Networking for per-destination load balancing



Router A needs to forward packets to 10.1.1.0/24 and 10.2.1.0/24. Based on per-destination load balancing, packets of the same flow are transmitted along the same path. The processes for forwarding packets on Router A are as follows:

- The first packet P1 to 10.1.1.0/24 is forwarded through Port 1, and all subsequent packets to 10.1.1.0/24 are forwarded through Port 1.
- The first packet P1 to 10.2.1.0/24 is forwarded through Port 2, and all subsequent packets to 10.2.1.0/24 are forwarded through Port 2.

Currently, RIP, OSPF, BGP, and IS-IS support load balancing, and static routes also support load balancing.

NOTE

The number of equal-cost routes for load balancing varies with the router model.

Route Backup

The MA5800 supports route backup to improve network reliability. You can configure multiple routes to the same destination as required. The route with the highest priority functions as the primary route, and the other routes with lower priorities function as backup routes.

In most cases, the router uses the primary route to forward packets. If the link fails, the primary route becomes inactive, and the router selects the backup route with the highest priority to forward data. In this way, traffic is switched to the backup route. When the original primary route recovers, the router restores and re-selects the optimal route. Because the original primary route has the highest priority, the router selects this route to send packets. In this way, the backup route is switched back to the primary route.

11.12.1.7 Re-advertisement of Routing Information

Different routing protocols may discover different routes because they adopt different routing algorithms. When the scale of a network is large and multiple

 Copyright © Huawei Technologies Co., Ltd.

routing protocols run on the network, these protocols need to re-advertise their discovered routes.

On the MA5800, the routes discovered by a routing protocol can be imported into the routing table of another routing protocol. Each protocol has its mechanism to import routes. For details, see "Routing Policy."

11.12.1.8 Default Route

Default routes are special routes. In most cases, they are configured by administrators. Default routes can also be generated by dynamic routing protocols, such as OSPF and IS-IS.

Default routes are used only when no matching routing entry is available for packet forwarding in the routing table. A default route in the routing table is the route to the network 0.0.0.0 (with mask 0.0.0.0). You can check whether the default route is configured using the **display ip routing-table** command.

If the destination address of a packet does not match any entry in the routing table, the packet is sent along a default route. If no default route exists and the destination address of the packet does not match any entry in the routing table, the packet is discarded. An Internet Control Message Protocol (ICMP) packet is then sent, informing the originating host that the destination host or network is unreachable.

11.12.1.9 IP FRR

Definition

IP fast reroute (IP FRR) provides fast traffic recovery upon link-layer or physical-layer faults. When a fault is detected and reported to the upper-layer routing system, a backup link is used to forward the impacted traffic.

IP FRR recovers the traffic impacted by link faults.

Purpose

If a forwarding link fails on a traditional IP network, the device port goes down. After detecting such a fault, the device notifies the upper-layer routing system so that it updates the link status and recalculates a route. The process takes several seconds.

The convergence time cannot meet the requirements of delay-sensitive services, such as voice over IP (VoIP) and online video services. In contrast, IP FRR can recover impacted traffic in the level of 50 ms.

Principle

IP FRR supports static routes and dynamic routes, such as OSPF, OSPFv3, IS-IS, IS-IS IPv6, BGP, and BGP4+ routes.

IP FRR is implemented as follows:

1. Run commands to configure and enable IP FRR.

 Copyright © Huawei Technologies Co., Ltd.

2. After IP FRR is enabled, the MA5800-X17 selects primary routes using routing protocols, calculates the backup links in advance, and saves them in the forwarding table.

3. If a primary link is unavailable, the MA5800-X17 forwards traffic through the backup link.

NOTE

For detailed configuration of IP FRR for each routing protocol, see the corresponding protocol description.

IP FRR Triggering Mode

IP FRR can be triggered by a BFD session or port status.

IP FRR triggered by a BFD session

If a Bidirectional Forwarding Detection (BFD) session is configured, it automatically associates with IP FRR. If a BFD session detects a link failure, the BFD session goes down and triggers IP FRR.

If multiple BFD sessions are configured and they have identical outgoing interfaces and destination IP addresses, IP FRR can be triggered if any of the BFD sessions goes down.

IP FRR triggered by port status

NOTE

This function is supported in V100R019C1X and later versions.

You can run the **port frr-binding** command to enable the function of triggering IP FRR by port status. When a port goes down, IP FRR is triggered.

The following restrictions apply when enabling the function of triggering IP FRR by port status:

- Virtual service ports do not support IP FRR. Only standard VLAN ports support IP FRR.
- To enable the function of triggering IP FRR by port status in a VLAN, the VLAN can have only one standard port, which can be a common Ethernet port or a port aggregation group.
- Protection groups do not support IP FRR.

11.12.1.10 ECMP

Equal and Weighted Cost Multi-Path (ECMP) is a technique in which if two or more equal cost shortest paths exist between two nodes, the traffic between the nodes is distributed among the multiple equal-cost paths.

11.12.1.10.1 Introduction to ECMP

Definition

Equal and Weighted Cost Multi-Path (ECMP) is a technique in which if two or more equal cost shortest paths exist between two nodes, the traffic between the

 Copyright © Huawei Technologies Co., Ltd.

nodes is distributed among the multiple equal-cost paths. That is, in packet transmissions, if different routes with the same destination network exist in the system, the packets can be transmitted to the destination network through multiple next hops.

Purpose

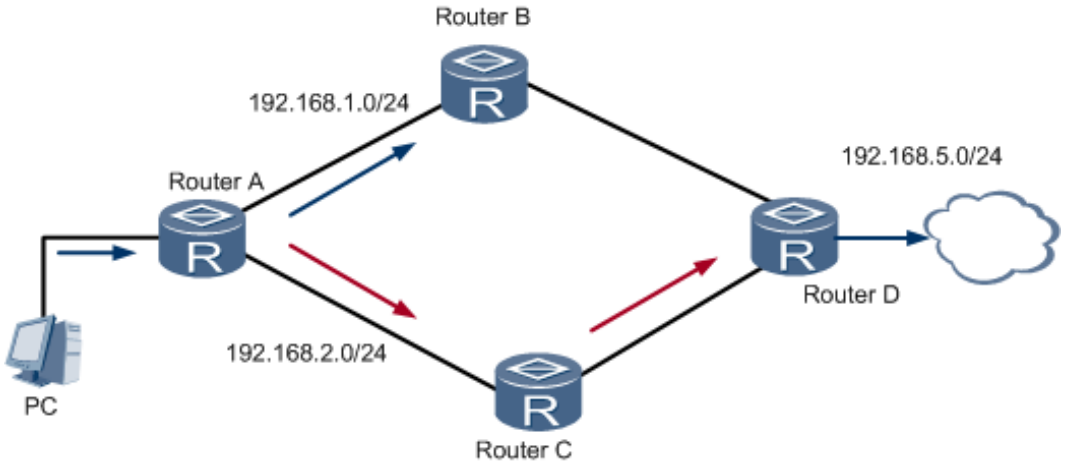
In ECMP, the traffic to the same destination network can be distributed among multiple equal-cost paths to reduce the network load, and the links in the network can back up each other. That is, when a link in the network fails, the packets on this link can be forwarded to the destination network through other links that are in the normal state.

11.12.1.10.2 ECMP Principle

In ECMP, according to different states of the network, the traffic to the same destination network can be distributed among multiple equal-cost paths to reduce the network load or to implement the link backup function.

As shown in **Figure 11-65**, assume that a packet is transmitted to the destination network (192.168.5.0) through Routers A-D, and two routes to the destination network exist in Router A. When receiving the packet from a user, Router A can select Router B or Router C as the next hop according to Hash algorithm to forward the packet to the destination network in load balancing mode.

Figure 11-65 ECMP diagram



The device supports two route load balancing modes. The **route load-balancing mode** command is used to configure the route load balancing mode.

- Based on the Hash algorithm of source and destination IP addresses, which is the default mode.
- Based on the Hash algorithm of IP address and UDP/TCP port.

11.12.1.10.3 ECMP Fast Switching

 Copyright © Huawei Technologies Co., Ltd.

Definition

The equal and weighted cost multi-path (ECMP) function transmits packets to the destination network through multiple paths of an equal cost. If one or multiple paths fail, the MA5800-X17 quickly switches traffic to normal paths. This process is known as ECMP fast switching.

Purpose

If one of the equal-cost paths fails, ECMP fast switching quickly switches the impacted traffic to another path to minimize traffic interruption duration.

Implementation

ECMP fast switching supports static routes and dynamic routes, such as OSPF, OSPFv3, IS-IS, IS-IS IPv6, BGP, and BGP4+ routes. ECMP fast switching can be triggered by a BFD session or port status change.

NOTE

If one of the equal-cost paths fails, the MA5800-X17 switches the impacted traffic to the adjacent path instead of evenly distributing the traffic among remaining available paths.

ECMP fast switching triggered by a BFD session

When a BFD session is configured, the session is automatically associated with ECMP fast switching. If a BFD session detects a link failure, the BFD session goes down and triggers ECMP fast switching.

ECMP fast switching triggered by port status

NOTE

This function is supported in V100R019C1X and later versions.

You can run the **port frr-binding** command to enable the function of triggering ECMP fast switching by port status. When a port failure is detected, ECMP fast switching is triggered.

The following restrictions apply when enabling the function of triggering ECMP fast switching by port status:

- Virtual service ports do not support ECMP fast switching. Only standard VLAN ports support ECMP fast switching.
- A VLAN with the function of triggering ECMP fast switching by port status enabled can have only one standard port, which can be a common Ethernet port or a port aggregation group.
- Protection groups do not support ECMP fast switching.

11.12.2 Static Route

 Copyright © Huawei Technologies Co., Ltd.

11.12.2.1 Introduction

Definition

Static routes are special routes that are configured by network administrators.

The MA5800 supports common static routes and the static routes associated with VPN instances. The static routes associated with VPN instances are used to manage VPN routes.

The MA5800 supports IPv4 static routes and IPv6 static routes.

Similar to IPv4 static routes, IPv6 static routes are configured by the administrator and are applicable to simple IPv6 networks. The major difference between IPv6 static routes and IPv4 static routes lies in their destination addresses and next hop addresses.

An IPv6 static route with destination address ::/0 (mask length 0) is a default IPv6 route. If the destination address of an IPv6 packet fails to match any entry in the routing table, a router selects the default IPv6 route to forward the IPv6 packet.

Purpose

On a simple network, only static routes can ensure that the network runs properly. Properly configuring and using static routes can improve network performance and guarantee the required bandwidth for important applications.

When a network fault occurs or the network topology changes, however, static routes must be changed manually by the administrator.

11.12.2.2 Components

On the MA5800, you can run the **ip route-static** command to configure a static route, which consists of the following components:

- **Destination address and mask**
- **Outbound interface and next hop address**

Destination Address and Mask

In the **ip route-static** command, the IPv4 address is expressed in dotted decimal notation, and the mask is expressed in dotted decimal notation or represented by the mask length (the number of consecutive 1s in the mask).

Outbound Interface and Next Hop Address

An outbound interface, a next hop address, or both of them can be configured for a static route.

Actually, each routing entry requires a next hop address. Before sending a packet, a device needs to search its routing table for the route matching the destination address in the packet based on the longest match rule. The device can find the associated link layer address to forward the packet only when the next hop address of the packet is available.

When specifying an outbound interface, note the following:

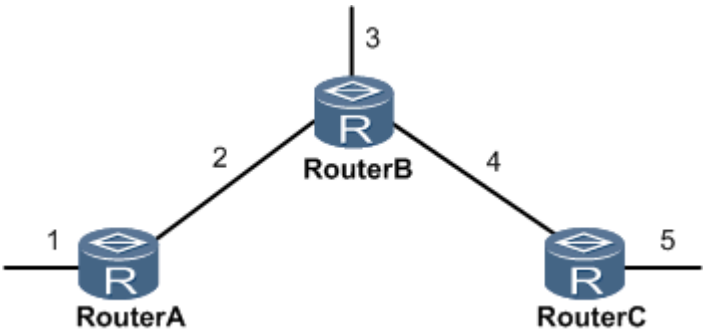
 Copyright © Huawei Technologies Co., Ltd.

- For a Point-to-Point (P2P) interface, if the outbound interface is specified, the next hop address is the address of the remote interface connected to the outbound interface. For example, when a POS interface is encapsulated with Point-to-Point Protocol (PPP) and obtains the remote IP address through PPP negotiation, you need to specify only the outbound interface rather than the next hop address.
- Non-Broadcast Multiple-Access (NBMA) interfaces are applicable to Point-to-Multipoint networks. Therefore, IP routes and the mappings between IP addresses and link layer addresses are required. In this case, you need to configure next hop addresses.
- An Ethernet interface is a broadcast interface and a virtual-template (VT) interface can be associated with multiple virtual access (VA) interfaces. If the Ethernet interface or the VT interface is specified as the outbound interface of a static route, the next hop cannot be determined because multiple next hops exist. Therefore, do not specify an Ethernet interface or a VT interface as the outbound interface unless necessary. If you need to specify a broadcast interface (such as an Ethernet interface) or a VT interface as the outbound interface, specify the associated next hop address.

11.12.2.3 Applications

As shown in **Figure 11-66**, the network topology is simple, and network communication can be implemented through static routes. You need to specify an address for each physical network, identify indirectly connected physical networks for each router, and configure static routes for indirectly connected physical networks.

Figure 11-66 Networking for static routes



In **Figure 11-66**, static routes to networks 3, 4, and 5 need to be configured on Router A; static routes to networks 1 and 5 need to be configured on Router B; static routes to networks 1, 2, and 3 need to be configured on Router C.

Default Static Route

The static route with the destination address and mask 0s (0.0.0.0 0.0.0.0) configured using the **ip route-static** command is a default route intended to simplify network configuration.

In **Figure 11-66**, because the next hop of the packets from Router A to networks 3, 4, and 5 is Router B, a default route can be configured on Router A to replace the three static routes destined for networks 3, 4, and 5. Similarly, only a default

 Copyright © Huawei Technologies Co., Ltd.

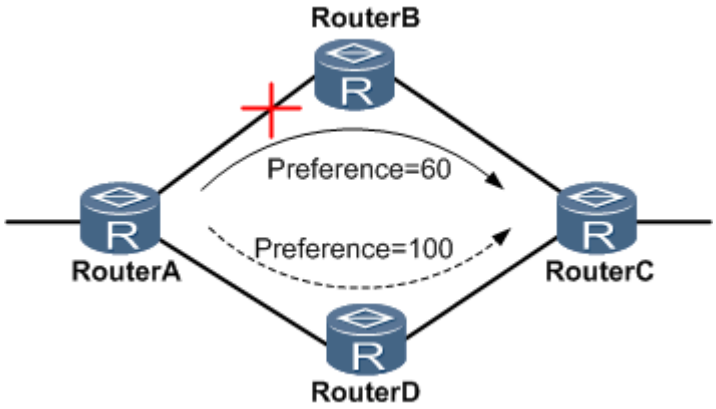
route from Router C to Router B needs to be configured to replace the three static routes destined for networks 1, 2, and 3.

Floating Static Routes

Different static routes can be configured with different priorities so that routing management policies can be flexibly applied. Route backup can be implemented by specifying different priorities for multiple routes to the same destination.

As shown in **Figure 11-67**, there are two static routes from Router A to Router C. In most cases, the only Active route is the static route with Router B as the next hop in the routing table because it has a higher priority. The other static route with Router D as the next hop functions as a backup route. The backup route is only activated to forward traffic when the primary link fails. After the primary link recovers, the static route with Router B as the next hop becomes Active to take over the traffic. Therefore, the backup route is also called a floating static route. The floating static route becomes ineffective if a fault occurs on the link between Router B and Router C.

Figure 11-67 Networking for a floating static route



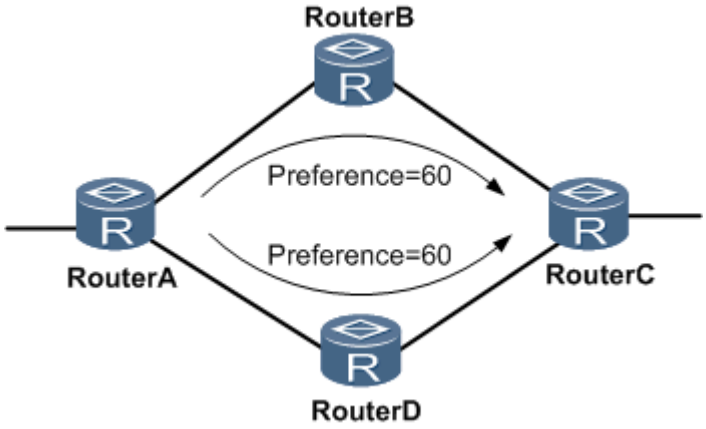
Load Balancing Among Static Routes

Routes to the same destination with the same priority can be used to load-balance traffic.

As shown in **Figure 11-68**, there are two static routes with the same priority from Router A to Router C. The two routes both exist in the routing table and forward traffic at the same time.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-68 Load balancing among static routes



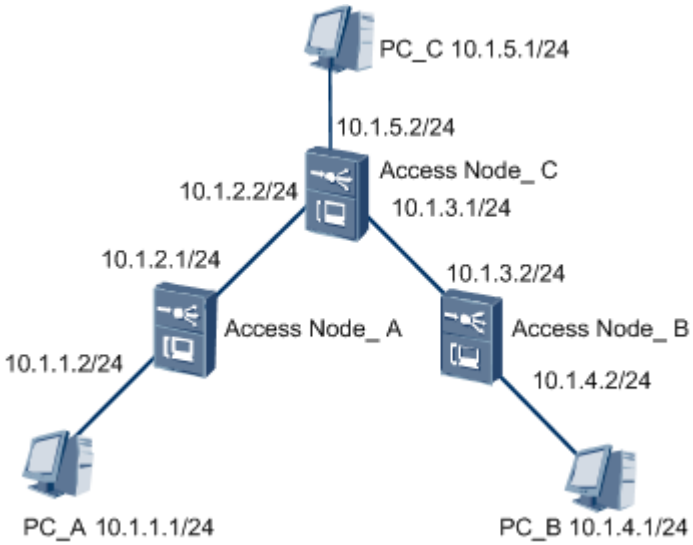
11.12.2.4 Configuration Example of the IPv4 Static Route

This topic describes how to manually add the IPv4 static route to implement the interconnection between MA5800.

Service Requirements

In this example network, MA5800_A, MA5800_B, and MA5800_C have the routing function. It is expected that after the IPv4 static route is configured, any two PCs can communicate with each other.

Figure 11-69 Example network for configuring the IPv4 static route



Procedure

Step 1 Configure the IP address of the Layer 3 interface.

The configurations for the three MA5800 devices are the same. The configuration of the MA5800 is considered as an example.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#vlan 2 smart
huawei(config)#port vlan 2 0/9 0
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.1.1.2 24
huawei(config-if-vlanif2)#ip address 10.1.2.1 24 sub
huawei(config-if-vlanif2)#quit
```

Step 2 Configure IPv4 static routes.

1. Configure an IPv4 static route for MA5800_A.

   ```
   huawei(config)#ip route-static 10.1.5.0 255.255.255.0 10.1.2.2
   huawei(config)#ip route-static 10.1.4.0 255.255.255.0 10.1.2.2
   ```

2. Configure an IPv4 static route for MA5800_B.

   ```
   huawei(config)#ip route-static 10.1.5.0 255.255.255.0 10.1.3.1
   huawei(config)#ip route-static 10.1.1.0 255.255.255.0 10.1.3.1
   ```

3. Configure IPv4 static routes for MA5800_C.

   ```
   huawei(config)#ip route-static 10.1.1.0 255.255.255.0 10.1.2.1
   huawei(config)#ip route-static 10.1.4.0 255.255.255.0 10.1.3.2
   ```

Step 3 Configure the host gateways.

1. Configure the default gateway of Host A to 10.1.1.2.
2. Configure the default gateway of Host B to 10.1.4.2.
3. Configure the default gateway of Host C to 10.1.5.2.

Step 4 Save the data.

```
huawei(config)#save
```

----End

Result

After the configuration, an interconnection can be set up between all the hosts and between all the MA5800 devices. Run the **ping** and **tracert** command to check the network connectivity.

Run the **display ip routing-table** command to query the IPv4 routing table which contains the static routing information that is configured.

Configuration File

Configuration example of MA5800_A.

```
vlan 2 smart
port vlan 2 0/9 0
interface vlanif 2
ip address 10.1.1.2 24
ip address 10.1.2.1 24 sub
quit
ip route-static 10.1.5.0 255.255.255.0 10.1.2.2
ip route-static 10.1.4.0 255.255.255.0 10.1.2.2
```

11.12.2.5 Configuration Example of the IPv6 Static Route

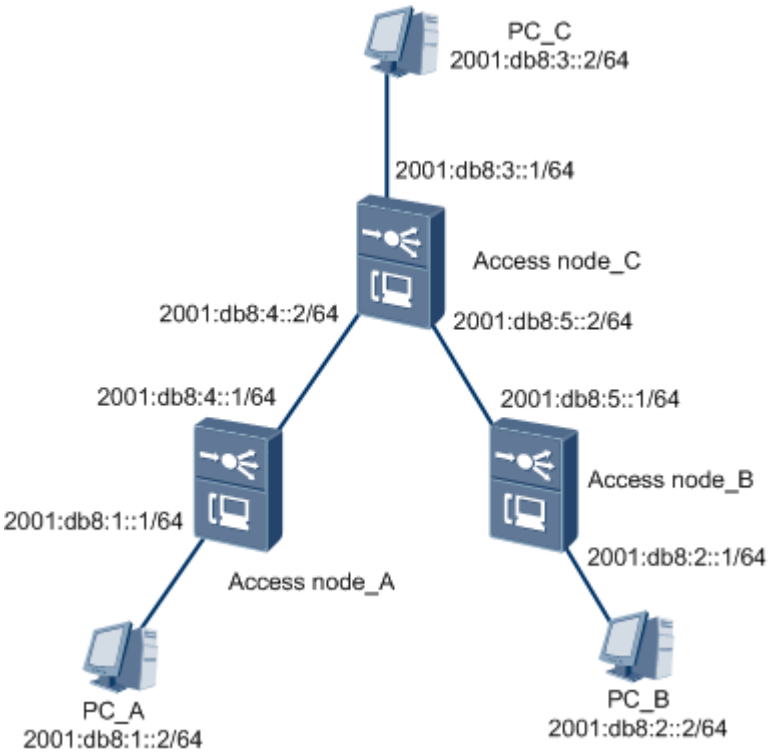
This topic describes how to manually add the IPv6 static route to implement the interconnection between access devices.

 Copyright © Huawei Technologies Co., Ltd.

Service Requirements

In this example network, three access nodes have the routing function. It is expected that after the IPv6 static route is configured, any two PCs can communicate with each other.

Figure 11-70 Example network for configuring the IPv6 static route



Procedure

Step 1 Configure the IPv6 address of the Layer 3 interface.

The configurations for the three access devices are the same. The configuration of the Access node_A is considered as an example.

```
huawei(config)#vlan 2 smart
huawei(config)#port vlan 2 0/9 0
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ipv6 enable
huawei(config-if-vlanif2)#ipv6 address 2001:db8:1::1 64
huawei(config-if-vlanif2)#quit
huawei(config)#vlan 3 smart
huawei(config)#port vlan 3 0/9 1
huawei(config)#interface vlanif 3
huawei(config-if-vlanif3)#ipv6 enable
huawei(config-if-vlanif3)#ipv6 address 2001:db8:4::1 64
huawei(config-if-vlanif3)#quit
```

Step 2 Configure IPv6 static routes.

1. Configure IPv6 static route for Access node_A.

   ```
   huawei(config)#ipv6 route-static :: 0 2001:db8:4::2
   ```

 Copyright © Huawei Technologies Co., Ltd.

2. Configure IPv6 static route for Access node_B.

```
huawei(config)#ipv6 route-static :: 0 2001:db8:5::2
```

3. Configure IPv6 static route for Access node_C.

```
huawei(config)#ipv6 route-static 2001:db8:1::1 64 2001:db8:4::1
huawei(config)#ipv6 route-static 2001:db8:2::1 64 2001:db8:5::1
```

Step 3 Configure the host gateways.

1. Configure the default gateway of Host A to 2001:db8:1::1.
2. Configure the default gateway of Host B to 2001:db8:2::1.
3. Configure the default gateway of Host C to 2001:db8:3::1.

Step 4 Save the data.

```
huawei#save
```

----End

Result

After the configuration, an interconnection can be set up between all the hosts and between all the access devices. Run the **ping ipv6** and **tracert ipv6** commands to query the network connectivity.

Run the **display ipv6 routing-table** command to query the IPv6 routing table which contains the static routing information that is configured.

Configuration File

Configuration example of Access node_A.

```
vlan 2 smart
port vlan 2 0/9 0
interface vlanif 2
ipv6 enable
ipv6 address 2001:db8:1::1/64
quit
vlan 3 smart
port vlan 3 0/9 1
interface vlanif 3
ipv6 enable
ipv6 address 2001:db8:4::1/64
quit
ipv6 route-static :: 0 2001:db8:4::2
```

11.12.2.6 Example for Configuring IP FRR for an IPv4 Static Route

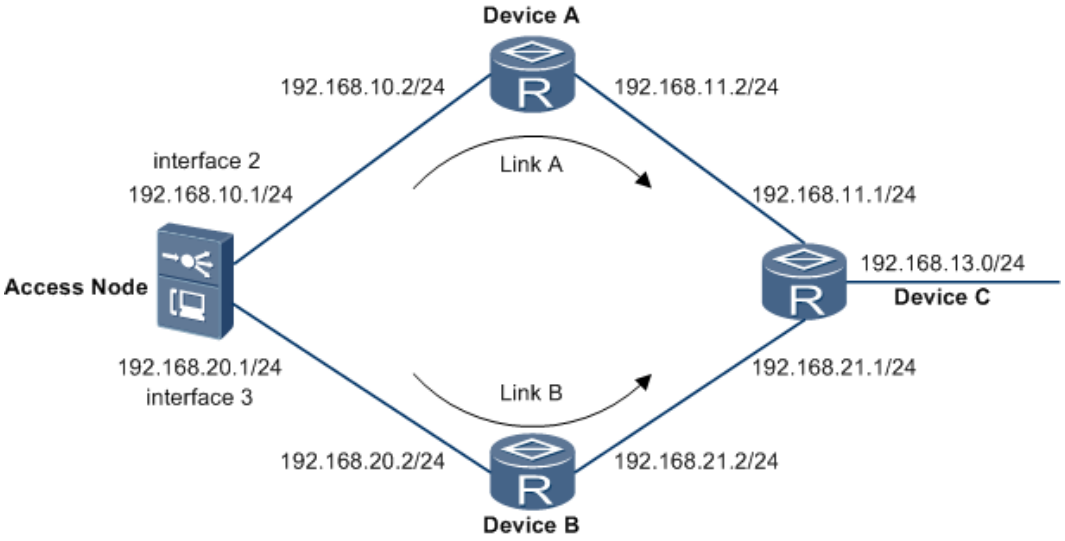
IP fast reroute (IP FRR) is applicable to services that are sensitive to packet loss and latency. Configuring IP FRR for IPv4 static routes implements link fast switching and relieves impacts on carried services brought by link faults.

Service Requirements

As shown in **Figure 11-71**, 2 IPv4 static routes having next hops device A and device B respectively are configured on the MA5800 so that link B is the backup for link A. In this case, when link A fails, traffic can be fast switched to link B.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-71 Networking of configuring IP FRR for an IPv4 static route



Data Plan

Table 11-18 describes data plan for configuring IP FRR for an IPv4 static route.

Table 11-18 Data plan for configuring IP FRR for an IPv4 static route

Item	Data
MA5800 port 2	IP address of the port: 192.168.10.1/24
	VLAN ID: 2
	Priority: 10
MA5800 port 3	IP address of the port: 192.168.20.1/24
	VLAN ID: 3
	Priority: default value

Prerequisite

Before configuring IP FRR for IPv4 static routes, complete the following task:

- You are advised to configure dynamic BFD for IPv4 static routes or static BFD for IPv4 static routes to quickly detect faults.

Procedure

Step 1 Perform configuration on the MA5800.

1. Configure the IP address of the Layer 3 interface.

```
huawei(config)#vlan 2 smart
huawei(config)#port vlan 2 0/9 0
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 192.168.10.1 24
huawei(config-if-vlanif2)#quit
huawei(config)#vlan 3 smart
huawei(config)#port vlan 3 0/9 1
huawei(config)#interface vlanif 3
huawei(config-if-vlanif3)#ip address 192.168.20.1 24
huawei(config-if-vlanif3)#quit
```

2. Configure an IPv4 static route and set a higher priority for link A so that the system preferentially selects link A.

```
huawei(config)#ip route-static 192.168.13.0 24 192.168.10.2 preference 10
huawei(config)#ip route-static 192.168.13.0 24 192.168.20.2
huawei(config)#ip route-static 192.168.11.0 24 192.168.10.2
huawei(config)#ip route-static 192.168.21.0 24 192.168.20.2
```

3. Enable IP FRR for an IPv4 static route.

```
huawei(config)#ip route-static frr
```

4. Save the data.

```
huawei(config)#save
```

----End

Result

Run the **display ip routing-table verbose** command to query the information about primary/secondary routes configured.

Configuration File

```
vlan 2 smart
port vlan 2 0/9 0
interface vlanif 2
ip address 192.168.10.1 24
quit
vlan 3 smart
port vlan 3 0/9 1
interface vlanif 3
ip address 192.168.20.1 24
quit
ip route-static 192.168.13.0 24 192.168.10.2 preference 10
ip route-static 192.168.13.0 24 192.168.20.2
ip route-static 192.168.11.0 24 192.168.10.2
ip route-static 192.168.21.0 24 192.168.20.2
ip route-static frr
save
```

11.12.2.7 Example for Configuring IP FRR for an IPv6 Static Route

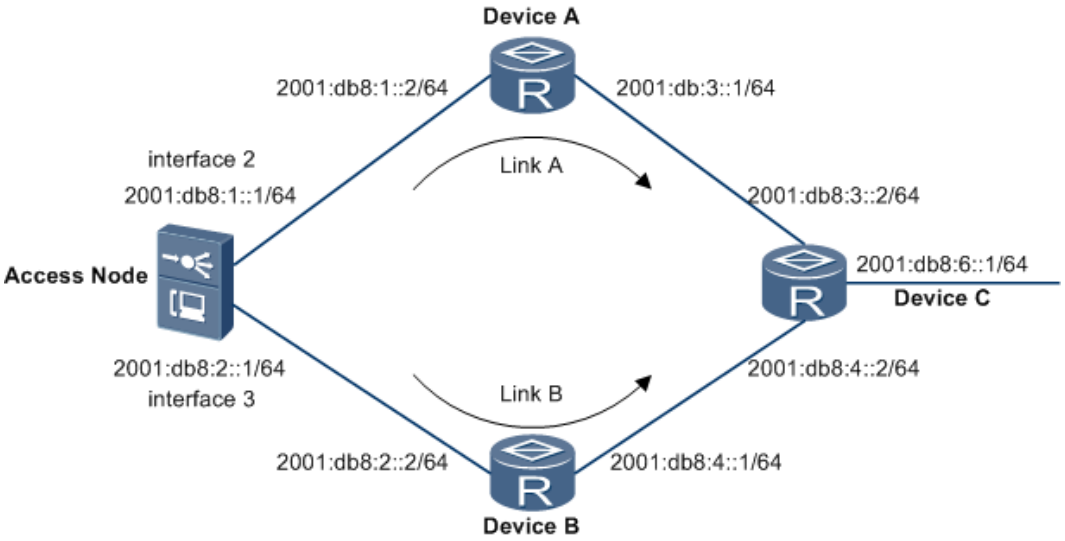
IP fast reroute (IP FRR) is applicable to services that are sensitive to packet loss and latency. Configuring IP FRR for IPv6 static routes implements link fast switching and relieves impacts on carried services brought by link faults.

Service Requirements

As shown in **Figure 11-72**, 2 IPv6 static routes having next hops device A and device B respectively are configured on the MA5800 so that link B is the backup for link A. In this case, when link A fails, traffic can be fast switched to link B.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-72 Networking of configuring IP FRR for an IPv6 static route



Data Plan

Table 11-19 describes data plan for configuring IP FRR for an IPv6 static route.

Table 11-19 Data plan for configuring IP FRR for an IPv6 static route

Item	Data
MA5800 port 2	IPv6 address of the port: 2001:db8:1::164
	VLAN ID: 2
	Priority: 10
MA5800 port 3	IPv6 address of the port: 2001:db8:2::164
	VLAN ID: 3
	Priority: default value

Prerequisite

Before configuring IP FRR for IPv6 static routes, complete the following task:

- You are advised to configure dynamic BFD for IPv6 static routes or static BFD for IPv6 static routes to quickly detect faults.

Procedure

Step 1 Perform configuration on the MA5800.

 Copyright © Huawei Technologies Co., Ltd.

1. Configure the IPv6 address of the Layer 3 interface.

```
huawei(config)#vlan 2 smart
huawei(config)#port vlan 2 0/9 0
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ipv6 enable
huawei(config-if-vlanif2)#ipv6 address 2001:db8:1::1 64
huawei(config-if-vlanif2)#quit
huawei(config)#vlan 3 smart
huawei(config)#port vlan 3 0/9 1
huawei(config)#interface vlanif 3
huawei(config-if-vlanif3)#ipv6 enable
huawei(config-if-vlanif3)#ipv6 address 2001:db8:2::1 64
huawei(config-if-vlanif3)#quit
```

2. Configure an IPv6 static route and set a higher priority for link A so that the system preferentially selects link A.

```
huawei(config)#ipv6 route-static 2001:db8:6::1 64 2001:db8:1::2 preference 10
huawei(config)#ipv6 route-static 2001:db8:6::1 64 2001:db8:2::2
huawei(config)#ipv6 route-static 2001:db8:3:: 64 2001:db8:1::2
huawei(config)#ipv6 route-static 2001:db8:4:: 64 2001:db8:2::2
```

3. Enable IPv6 FRR for an IPv4 static route.

```
huawei(config)#ipv6 route-static frr
```

4. Save the data.

```
huawei(config)#save
```

----End

Result

Run the **display ipv6 routing-table verbose** command to query the information about primary/secondary routes configured.

Configuration File

```
vlan 2 smart
port vlan 2 0/9 0
interface vlanif 2
ipv6 enable
ipv6 address 2001:db8:1::1 64
quit
vlan 3 smart
port vlan 3 0/9 1
interface vlanif 3
ipv6 enable
ipv6 address 2001:db8:2::1 64
quit
ipv6 route-static 2001:db8:6:: 64 2001:db8:1::2 preference 10
ipv6 route-static 2001:db8:6:: 64 2001:db8:2::2
ipv6 route-static 2001:db8:3:: 64 2001:db8:1::2
ipv6 route-static 2001:db8:4:: 64 2001:db8:2::2
ipv6 route-static frr
save
```

11.12.3 RIP

 Copyright © Huawei Technologies Co., Ltd.

11.12.3.1 Introduction

Definition

Routing Information Protocol (RIP) is a simple Interior Gateway Protocol (IGP). RIP is used in small-scale networks, such as campus networks and simple regional networks.

As a distance-vector routing protocol, RIP exchanges routing information through User Datagram Protocol (UDP) packets with port number 520.

RIP employs the hop count as the metric to measure the distance to the destination. In RIP, by default, the number of hops from the Router to its directly connected network is 0; the number of hops from the Router to a network that is reachable through another Router is 1, and so on. The hop count (the metric) equals the number of Routers along the path from the local network to the destination network. To speed up route convergence, RIP defines the hop count as an integer that ranges from 0 to 15. A hop count that is equal to or greater than 16 is classified as infinite, indicating that the destination network or host is unreachable. Due to the hop limit, RIP is not applicable to large-scale networks.

RIP supports split horizon, poison reverse, and triggered update, which improves the performance and prevents routing loops.

Purpose

As the earliest IGP, RIP is used in small and medium-sized networks. Its implementation is simple, and the configuration and maintenance of RIP are easier than those of Open Shortest Path First (OSPF) and Intermediate System-to-Intermediate System (IS-IS). Therefore, RIP is widely used on live networks.

11.12.3.2 Principles

RIP is a distance-vector routing protocol. It forwards packets through UDP and uses timers to control the advertisement, update, and aging of routing information. However, design defects in RIP may cause routing loops. Therefore, split horizon, poison reverse, and triggered update were introduced into RIP to prevent routing loops.

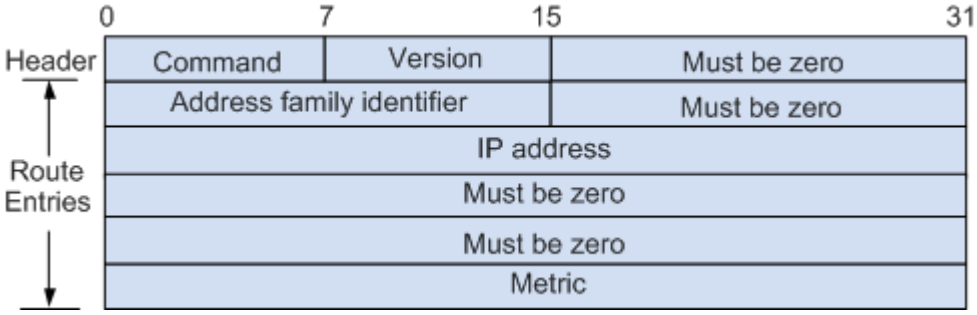
In addition, RIP periodically advertises its routing table to neighbors, and route summarization was introduced to reduce the size of the routing table.

11.12.3.2.1 RIP-1

RIP Version 1 (RIP-1) is a classful routing protocol, which supports only the broadcast of protocol packets. **Figure 11-73** shows the format of a RIP-1 packet. A RIP packet can carry a maximum of 25 routing entries. RIP is based on UDP, and a RIP-1 packet cannot be longer than 512 bytes. RIP-1 packets do not carry any mask information, and RIP-1 can identify only the routes to natural network segments, such as Class A, Class B, and Class C. Therefore, RIP-1 does not support route aggregation or discontinuous subnets.

 Copyright © Huawei Technologies Co., Ltd.

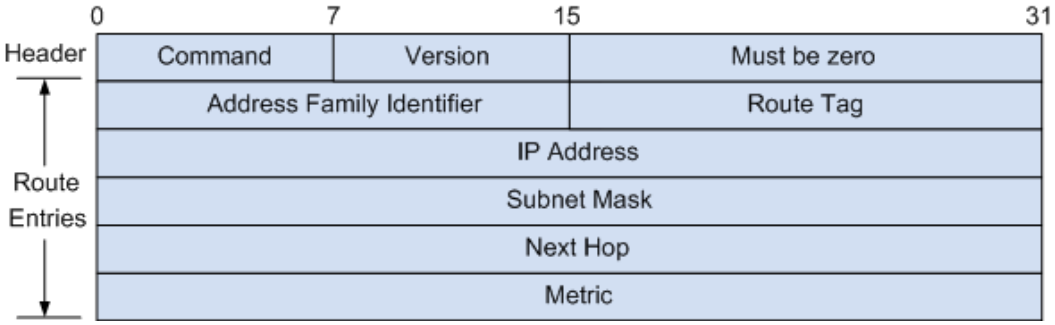
Figure 11-73 RIP-1 packet format



11.12.3.2.2 RIP-2

RIP Version 2 (RIP-2) is a classless routing protocol. **Figure 11-74** shows the format of a RIP-2 packet.

Figure 11-74 RIP-2 packet format



Compared with RIP-1, RIP-2 has the following advantages:

- Supports external route tags and flexibly controls routes based on the tag using a routing policy.
- Supports route aggregation and Classless Inter-domain Routing (CIDR) by adding mask information to RIP-2 packets.
- Supports next hop specification so that the optimal next hop address can be specified on the broadcast network.
- Supports Update packets transmission along multicast routes. Only the Routers that support RIP-2 can receive RIP-2 packets, which reduces resource consumption.
- Provides three packet authentication modes: simple authentication, Message Digest 5 (MD5) authentication and HMAC-MD5 authentication.

NOTICE

MD5 is an insecure encryption algorithm. Therefore, using HMAC-SHA256 is recommended.

11.12.3.2.3 Timers

RIP uses the following timers:

 Copyright © Huawei Technologies Co., Ltd.

- Update timer: The Update timer periodically triggers Update packet transmission. By default, the interval at which Update packets are sent is 30s.
- Age timer: If a RIP device does not receive any packets from its neighbor to update a route before the route expires, the RIP device considers the route unreachable. By default, the age timer interval is 180s.
- Garbage-collect timer: If a route becomes invalid after the age timer expires or a route unreachable message is received, the route is placed into a garbage queue instead of being immediately deleted from the RIP routing table. The garbage-collect timer monitors the garbage queue and deletes expired routes. If an Update packet of a route is received before the garbage-collect timer expires, the route is placed back into the age queue. The garbage-collect timer is set to avoid route flapping. By default, the garbage collect timer interval is 120s.
- Hold-down timer: If a RIP device receives an updated route with cost 16 from a neighbor, the route enters the holddown state, and the hold-down timer is started. To avoid route flapping, the RIP device does not accept any updated routes until the hold-down timer expires, even if the cost is less than 16 except in the following scenarios:
  a. The cost carried in the update packet is less than or equal to that carried in the last update packet.
  b. The hold-down timer expires, and the corresponding route enters the Garbage state.

The relationship between RIP routes and the four timers is as follows:

- The advertisement of RIP routing updates is triggered by the update timer with a default value 30 seconds.
- Each routing entry is associated with two timers: the age timer and garbage-collect timer.
  a. Each time a route is learned and added to the routing table, the age timer is started.
  b. If no Update packet is received from the neighbor within 180 seconds after the age timer is started, the metric of the corresponding route is set to 16, and the garbage-collect timer is started.
- If no Update packet is received within 120 seconds after the garbage-collect timer is started, the corresponding routing entry is deleted from the routing table after the garbage-collect timer expires.
- By default, the hold-down timer is disabled. If you configure a hold-down timer, it starts after the system receives a route with a cost greater than 16 from its neighbor.

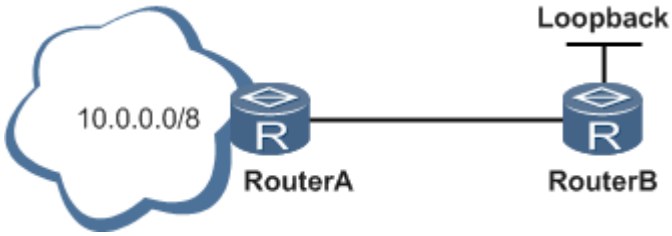
11.12.3.2.4 Split Horizon

Split Horizon on Broadcast, P2PMP, and P2P Networks

Split horizon prevents a RIP-enabled interface from sending back the routes it learns, which reduces bandwidth consumption and prevents routing loops.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-75 Networking for interface-based split horizon



In **Figure 11-75**, Router A sends Router B a route to 10.0.0.0/8. If split horizon is not configured, Router B will send this route back to Router A after learning it from Router A. As a result, Router A learns the following routes to 10.0.0.0/8:

- A direct route with zero hops
- A route with Router B as the next hop and total two hops

Only direct routes, however, are active in the RIP routing table of Router A.

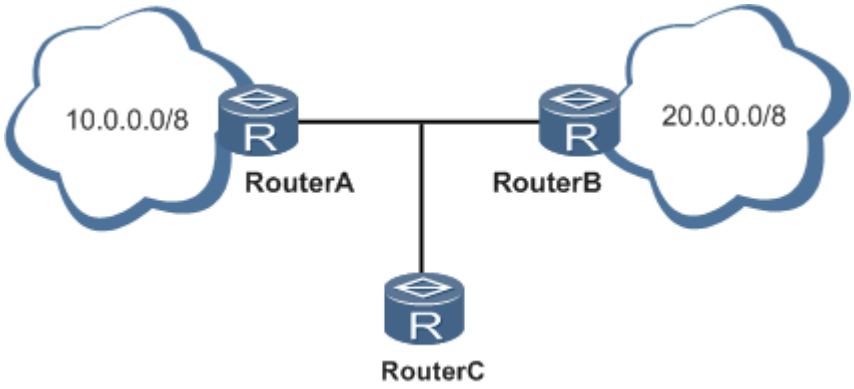
If the route from Router A to 10.0.0.0/8 becomes unreachable and Router B is not notified, Router B still considers the route to 10.0.0.0/8 reachable and continues sending this route to Router A. Then, Router A receives incorrect routing information and regards the route to 10.0.0.0/8 reachable through Router B; Router B considers the route to 10.0.0.0/8 reachable through Router A. As a result, a loop occurs on the network.

After split horizon is configured, Router B no longer sends the route back after learning the route, which prevents such a loop.

Split Horizon on NBMA Networks

On a Non-Broadcast Multi-Access (NBMA) network where an interface is connected to multiple neighbors, RIP supports neighbor-based split horizon. On NBMA networks, routes are sent in unicast mode, and an interface can differentiate which neighbor each route was learned from, and the interface will not send the routes back to the neighbor it learned them from.

Figure 11-76 Networking for neighbor-based split horizon on an NBMA network



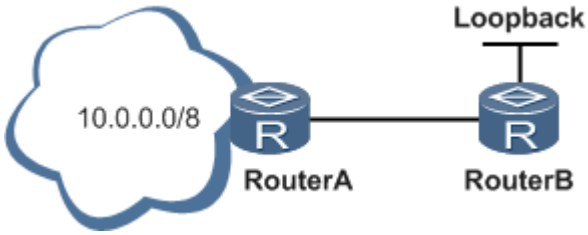
As shown in **Figure 11-76**, Router A sends the route to 10.0.0.0/8 that it learns from Router B only to Router C.

 Copyright © Huawei Technologies Co., Ltd.

11.12.3.2.5 Poison Reverse

Poison reverse allows a RIP-enabled interface to set the cost of the route that it learns from a neighbor to 16 (indicating that the route is unreachable) and then send the route back. After receiving this route, the neighbor deletes the useless route from its routing table, which prevents loops.

Figure 11-77 Networking for poison reverse



As shown in **Figure 11-77**, Router A sends Router B a route to 10.0.0.0/8. If poison reverse is not configured, Router B will send this route back to Router A after learning it from Router A. As a result, Router A learns the following routes to 10.0.0.0/8:

- A direct route with zero hops
- A route with Router B as the next hop and total two hops

Only direct routes, however, are active in the RIP routing table of Router A.

If the route from Router A to 10.0.0.0 becomes unreachable and Router B is not notified, Router B still considers the route to 10.0.0.0/8 reachable and continues sending this route to Router A. Then, Router A receives incorrect routing information and considers the route to 10.0.0.0/8 reachable through Router B; Router B considers the route to 10.0.0.0/8 reachable through Router A. As a result, a loop occurs on the network.

With poison reverse, after Router B receives the route from Router A, Router B sends a route unreachable message to Router A with cost 16. Router A then no longer learns the reachable route from Router B, which prevents routing loops.

If both split horizon and poison reverse are configured, only poison reverse takes effect.

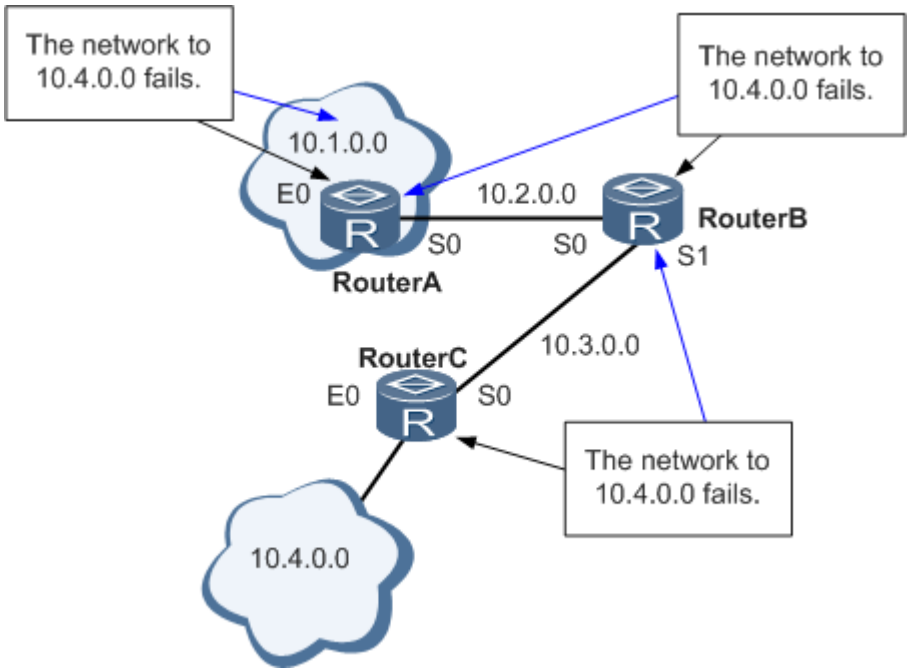
11.12.3.2.6 Triggered Update

Triggered update allows a device to advertise routing information changes to neighbor routers immediately.

Triggered update shortens the network convergence time. When local routing information changes, the local router immediately notifies its neighbor routers of the changes rather than waiting for periodical update.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-78 Networking for triggered update



In the networking shown in **Figure 11-78**, if the route to 10.4.0.0 becomes unreachable, Router C learns the information first. By default, a RIP-enabled device sends routing updates to its neighbors every 30s. If Router C receives an Update packet from Router B within 30s while Router C is still waiting to send Update packets, Router C learns the incorrect route to 10.4.0.0. In this case, the next hops of the routes from Router B or Router C to network 10.4.0.0 are Router C and Router B respectively, which results in routing loops. If Router C sends an Update packet to Router B immediately after it detects a network, Router B can rapidly update its routing table, which prevents routing loops.

In addition, if the next hop of a route becomes unavailable due to a link failure, the local Router sets the cost of the route to 16 and then advertises the route immediately to its neighbors. This process is called route poisoning.

11.12.3.2.7 Route Summarization

Route summarization allows routes to the same natural network segment but different subnets to be summarized into a single route belonging to the same network segment before it is transmitted to other network segments. RIP-1 packets do not carry mask information, and therefore RIP-1 can advertise only routes with natural masks. RIP-2 supports route summarization because RIP-2 packets carry mask information. Therefore, RIP-2 supports subnetting.

In RIP-2, route summarization can reduce the size of the routing table and improve the extensibility and efficiency of a large-scale network.

Route summarization has two modes:

- Process-based classful summarization

  Summarized routes are advertised with natural masks. If split horizon or poison reverse is configured, classful summarization becomes invalid because split horizon or poison reverse suppresses some routes from being advertised.

 Copyright © Huawei Technologies Co., Ltd.

In addition, when classful summarization is configured, routes learned from different interfaces may be summarized into a single route. As a result, a conflict occurs in the advertisement of the summarized route.

For example, a RIP process summarizes the route 10.1.1.0 /24 with metric 2 and route 10.2.2.0/24 with metric 3 into the route 10.0.0.0/8 with metric 2.

- Interface-based summarization

  Users can specify a summary address.

  For example, users can configure a RIP-enabled interface to summarize the route 10.1.1.0/24 with metric 2 and route 10.2.2.0/24 with metric 3 into the route 10.1.0.0/16 with metric 2.

11.12.3.2.8 Multi-Process

RIP supports multi-process and multi-instance to simplify network management and improve service control efficiency. Multi-process allows a set of interfaces to be associated with a specific RIP process, which ensures that the specific RIP process performs all the protocol operations only on this set of interfaces. Therefore, multiple RIP processes can run on one Router, and each process manages a unique set of interfaces. In addition, the routing data of each RIP process is independent; however, processes can import routes from each other.

11.12.3.2.9 RIP NSR

Devices with a distributed architecture support RIP Non-stop Routing (NSR). RIP backs up all route data from the Active Main Board (AMB) to the Standby Main Board (SMB). Whenever the AMB fails, the SMB becomes active and takes over traffic. RIP NSR ensures that routes are synchronous between the AMB and SMB. Therefore, during the AMB/SMB switchover, the neighbor will not detect the fault on the local device.

11.12.3.2.10 RIP Authentication

As networks develop, there has been considerable growth in all types of data, voice, and video information exchanged on networks. In addition, new services, such as E-commerce, online conferencing and auctions, video on demand (VoD), and e-learning have sprung up increasingly, requiring higher information security than before. Carriers must protect data packets from being intercepted or modified by attackers and prohibit unauthorized users from accessing network resources. RIP packet authentication effectively meets these security requirements.

RIP authentication falls into the following modes:

- Simple authentication: The authenticated party adds the configured password directly to packets for authentication. This authentication mode provides the lowest password security.
- MD5 authentication: The authenticated party uses the Message Digest 5 (MD5) algorithm to generate a ciphertext password and adds it to packets for authentication. This authentication mode improves password security.

NOTICE

MD5 is an insecure encryption algorithm. Therefore, using HMAC-SHA256 is recommended.

 Copyright © Huawei Technologies Co., Ltd.

- HMAC-SHA256 authentication: The authenticated party uses the HMAC-SHA256 algorithm to generate a ciphertext password and adds it to packets for authentication.

RIP authentication ensures network security by adding an authentication field used to encrypt a packet before sending the packet to ensure network security. After receiving a RIP packet from a remote router, the local router discards the packet if the authentication password in the packet does not match the local authentication password. This authentication mode protects the local router.

On IP networks of carriers, RIP authentication ensures the secure transmission of packets, improves the system security, and provides secure network services for carriers.

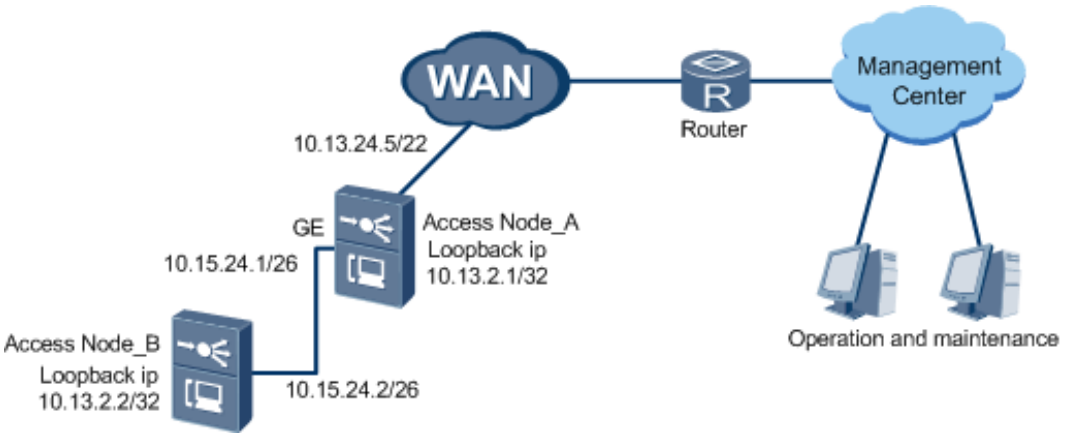
11.12.3.3 Configuration Example of RIP

This topic provides an example for configuring RIP on the MA5800.

Service Requirements

- MA5800_A is cascaded with MA5800_B through port 0/9/1, and uses port 0/9/0 to transmit services in the upstream. Besides, it connects to the management center network through the WAN.
- RIP is enabled on MA5800_A and MA5800_B so that the administrator can access MA5800_A and MA5800_B through the RIP route. Then, you can operate and maintain MA5800_A and MA5800_B.

Figure 11-79 Example network for configuring RIP



Data Plan

Table 11-20 provides the data plan for configuring RIP.

 Copyright © Huawei Technologies Co., Ltd.

Table 11-20 Data plan for configuring RIP

Item	Data
MA5800_A	Upstream port: 0/9/0 Administration VLAN: smart VLAN 100 IP address of the Layer 3 interface in the administration VLAN: 10.13.24.5/22 Loopback interface address: 10.13.2.1/32
	RIP version: V2 RIP route filtering policy: filtering routes based on the IP address prefix list "abc". Only the routes with the IP addresses 10.13.2.1 and 10.13.2.2 can be advertised through the Layer 3 interface of VLAN 100.
	Cascading port: 0/9/0 Cascading administration VLAN: smart VLAN 10 IP address of the Layer 3 interface in the cascading administration VLAN: 10.15.24.1/26
MA5800_B	Cascading port: 0/9/1 Administration VLAN: smart VLAN 10 IP address of the Layer 3 interface in the administration VLAN: 10.15.24.2/26 Loopback interface address: 10.13.2.2/32
	RIP version: V2 RIP route filtering policy: filtering routes based on the IP address prefix list "abc". Only the route with the IP address 10.13.2.2 can be advertised through the Layer 3 interface of VLAN 10.

Procedure

- Configure MA5800_A.
  a. Configure the RIP-supported Layer 3 interface.

```
huawei(config)#vlan 100 smart
huawei(config)#port vlan 100 0/9 0
huawei(config)#interface vlanif 100
huawei(config-if-vlanif100)#ip address 10.13.24.5 22
huawei(config-if-vlanif100)#quit
huawei(config)#interface loopBack 0
huawei(config-if-loopback0)#ip address 10.13.2.1 32
huawei(config-if-loopback0)#quit
```

  b. Enable RIP.

```
huawei(config)#rip 1
huawei(config-rip-1)#network 10.0.0.0
huawei(config-rip-1)#version 2
huawei(config-rip-1)#quit
```

  c. Configure the route filtering policy.

```
huawei(config)#ip ip-prefix abc permit 10.13.2.1 32
huawei(config)#ip ip-prefix abc permit 10.13.2.2 32
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#rip 1
huawei(config-rip-1)#filter-policy ip-prefix abc export vlanif 100
huawei(config-rip-1)#quit
```

d. Configure the cascading port.

```
huawei(config)#vlan 10 smart
huawei(config)#port vlan 10 0/9 1
huawei(config)#interface mpu 0/9
huawei(config-if-mpu-0/9)#network-role 1 cascade
huawei(config-if-mpu-0/9)#quit
huawei(config)#interface vlanif 10
huawei(config-if-vlanif10)#ip address 10.15.24.1 26
huawei(config-if-vlanif10)#quit
```

e. Enable RIP on the cascading port.

```
huawei(config)#rip 1
huawei(config-rip-1)#network 10.0.0.0
huawei(config-rip-1)#quit
```

f. Save the data.

```
huawei(config)#save
```

- Configure MA5800_B.

a. Configure the RIP-supported Layer 3 interface.

```
huawei(config)#vlan 10 smart
huawei(config)#port vlan 10 0/9 0
huawei(config)#interface vlanif 10
huawei(config-if-vlanif10)#ip address 10.15.24.2 26
huawei(config-if-vlanif10)#quit
huawei(config)#interface loopBack 0
huawei(config-if-loopback0)#ip address 10.13.2.2 32
huawei(config-if-loopback0)#quit
```

b. Enable RIP.

```
huawei(config)#rip 1
huawei(config-rip-1)#network 10.0.0.0
huawei(config-rip-1)#version 2
huawei(config-rip-1)#quit
```

c. Configure the route filtering policy.

```
huawei(config)#ip ip-prefix abc permit 10.13.2.2 32
huawei(config)#rip 1
huawei(config-rip-1)#filter-policy ip-prefix abc export vlanif 10
huawei(config-rip-1)#quit
```

d. Save the data.

```
huawei(config)#save
```

----End

Result

The maintenance terminal of the administration center can access MA5800_A and MA5800_B, and operate and maintain the two devices.

Configuration File

Configuration on MA5800_A

```
vlan 100 smart
port vlan 100 0/9 0
interface vlanif 100
ip address 10.13.24.5 22
quit
interface loopBack 0
ip address 10.13.2.1 24
```

 Copyright © Huawei Technologies Co., Ltd.

```
quit
rip 1
network 10.0.0.0
version 2
quit
ip ip-prefix abc permit 10.13.2.1 32
ip ip-prefix abc permit 10.13.2.2 32
rip 1
filter-policy ip-prefix abc export vlanif 100
quit
vlan 10 smart
port vlan 10 0/9 1
interface mpu 0/9
network-role 1 cascade
quit
interface vlanif 10
ip address 10.15.24.1 26
quit
rip 1
network 10.0.0.0
quit
save
```

Configuration on MA5800_B

```
vlan 10 smart
port vlan 10 0/9 0
interface vlanif 10
ip address 10.15.24.2 26
quit
interface loopBack 0
ip address 10.13.2.2 32
quit
rip 1
network 10.0.0.0
version 2
quit
ip ip-prefix abc permit 10.13.2.2 32
rip 1
filter-policy ip-prefix abc export vlanif 10
quit
save
```

11.12.3.4 RIP Reference Standards and Protocols

The following table lists the references.

Document NO.	Document Name
RFC 1058	Routing Information Protocol
RFC 1721	RIP Version 2 Protocol Analysis
RFC 1722	RIP Version 2 Protocol Applicability Statement
RFC 1723	RIP Version 2 - Carrying Additional Information
RFC 2082	RIP-2 MD5 Authentication
RFC 2453	RIP Version 2

11.12.4 RIPng

 Copyright © Huawei Technologies Co., Ltd.

11.12.4.1 Introduction

Definition

RIP next generation (RIPng) is an extension of RIPv2 in IPv4 networks to support of IPv6, the next generation Internet Protocol. Most RIP concepts apply to RIPng.

RIPng, based on the Distance Vector (D-V) algorithm, is a routing protocol that measures the distance (metrics or cost) to the host by Hop Count (HC). RIPng defines that the HC from a device to its directly connected network is 0, and the HC from a device to a network that is reachable through another device is 1, and so on. When the HC reaches 16, the destination network or host is defined as unreachable.

To be applied on IPv6 networks, RIPng makes the following changes to RIP:

- RIPng uses UDP port 521 to send and receive routing information. RIPv2 uses UDP port 520.
- Multicast address: RIPng uses FF02::9 as the link-local multicast address of a RIPng device.
- Prefix length: RIPng uses a 128-bit (the mask length) prefix in the destination address.
- Next hop address: RIPng uses a 128-bit IPv6 address.
- Source address: RIPng uses link-local address FE80::/10 as the source address to send RIPng Update packets.

Purpose

RIPng is an extension to RIP for support of IPv6.

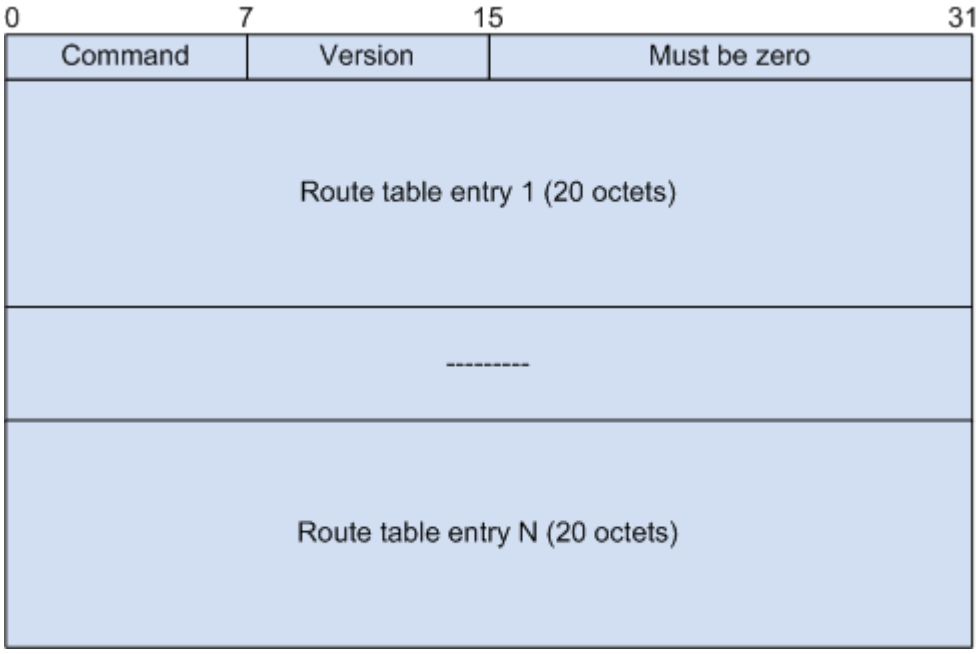
11.12.4.2 RIPng Packet Format

A RIPng packet is composed of a header and multiple route table entries (RTEs). In a RIPng packet, the maximum number of RTEs is determined by the maximum transmission unit (MTU) of an interface.

Figure 11-80 shows the basic format of a RIPng packet.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-80 RIPng packet format



A RIPng packet contains two types of RTEs:

- Next hop RTE: It defines the IPv6 address of the next hop and is located before a group of IPv6-prefix RTEs that have the same next hop. The Metric field of a next hop RTE is always 0xFF.
- IPv6-prefix RTE: It describes the destination IPv6 address and the cost in the RIPng routing table and is located after a next hop RTE. A next hop RTE can be followed by multiple different IPv6-prefix RTEs.

Figure 11-81 shows the format of a next hop RTE.

Figure 11-81 Format of the next hop RTE

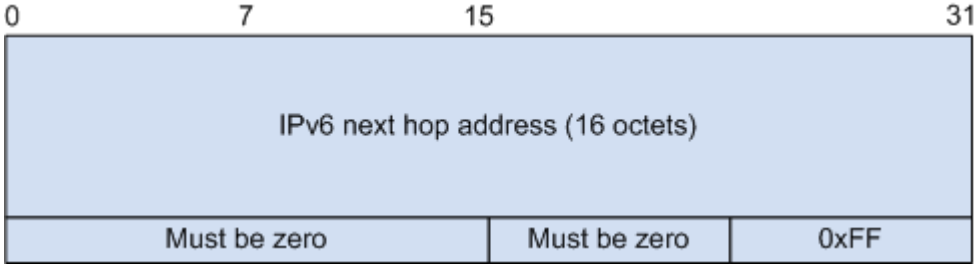
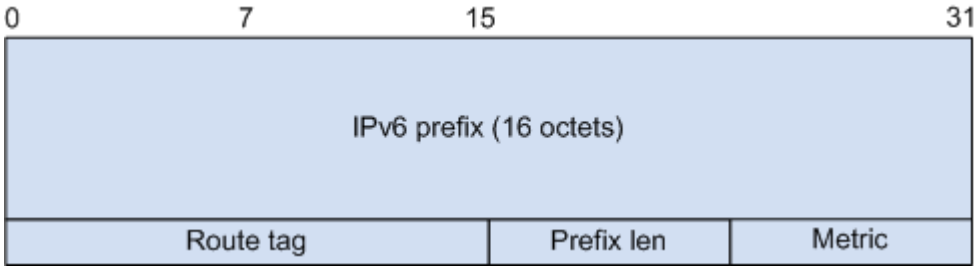


Figure 11-82 shows the format of an IPv6-prefix RTE.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-82 Format of the IPv6-prefix RTE



11.12.4.3 Timers

RIPng uses the following timers:

- Update timer: This timer periodically triggers Update packet transmission. By default, the interval at which Update packets are sent is 30s. This timer is used to synchronize RIPng routes on the network.
- Age timer: If a RIPng device does not receive any Update packet from its neighbor before a route expires, the RIPng device considers the route to its neighbor unreachable.
- Garbage-collect timer: If no packet is received to update an unreachable route after the Age timer expires, this route is deleted from the RIPng routing table.
- Hold-down timer: If a RIP device receives an updated route with cost 16 from a neighbor, the route enters the holddown state, and the hold-down timer is started.

The relationship between these timers is as follows:

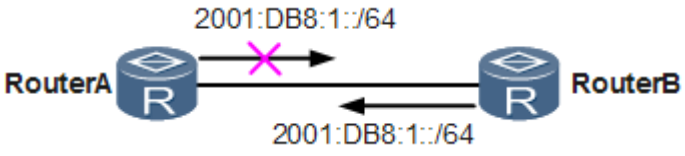
The advertisement of RIPng routing updates is periodically triggered by the update timer with default value 30 seconds. Each routing entry is associated with the age timer and garbage-collect timer. Each time a route is learned and added to the routing table, the age timer is started. If no update packet is received from the neighbor within 180 seconds, the metric of the route is set to 16, and the garbage-collect timer is started. If no update packet is received within 120 seconds, the route is deleted after the garbage-collect timer expires.

By default, hold-down timer is disabled. If you configure a hold-down timer, it starts after the system receives a route with a cost greater than 16 from its neighbor.

11.12.4.4 Split Horizon

Split horizon prevents a RIPng-enabled interface from sending back the routes it learns, which reduces bandwidth consumption and prevents routing loops.

Figure 11-83 Networking for split horizon



 Copyright © Huawei Technologies Co., Ltd.

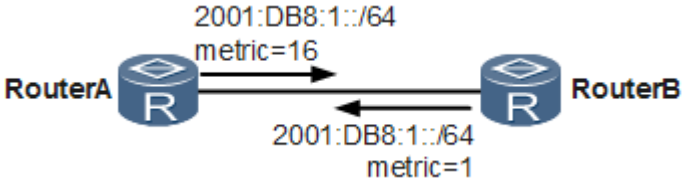
On the network shown in **Figure 11-83**, after Router B sends a route to network 2001:DB8:1::/64 to Router A, Router A does not send the route back to Router B.

11.12.4.5 Poison Reverse

Poison reverse allows a RIPng-enabled interface to set the cost of the route that it learns from a neighbor to 16 (indicating that the route is unreachable) and then send the route back. In this way, RIPng can delete useless routes from the routing table of the neighbor.

Poison reverse of RIPng can also avoid route loops.

Figure 11-84 Networking for poison reverse



On the network shown in **Figure 11-84**, if poison reverse is not configured, Router B sends Router A a route learned from Router A. The metric of the route from Router A to network 2001:DB8:1::/64 is 1. If the route from Router A to network 2001:DB8:1::/64 becomes unreachable and Router B does not receive an Update packet from Router A and keeps sending Router A the route from Router A to network 2001:DB8:1::/64, a routing loop occurs.

With poison reverse, after Router A receives the route from Router B, Router A sends a route unreachable message to Router B with metric 16. Router B then no longer learns the reachable route from Router A, which prevents routing loops.

If both poison reverse and split horizon are configured, only poison reverse takes effect.

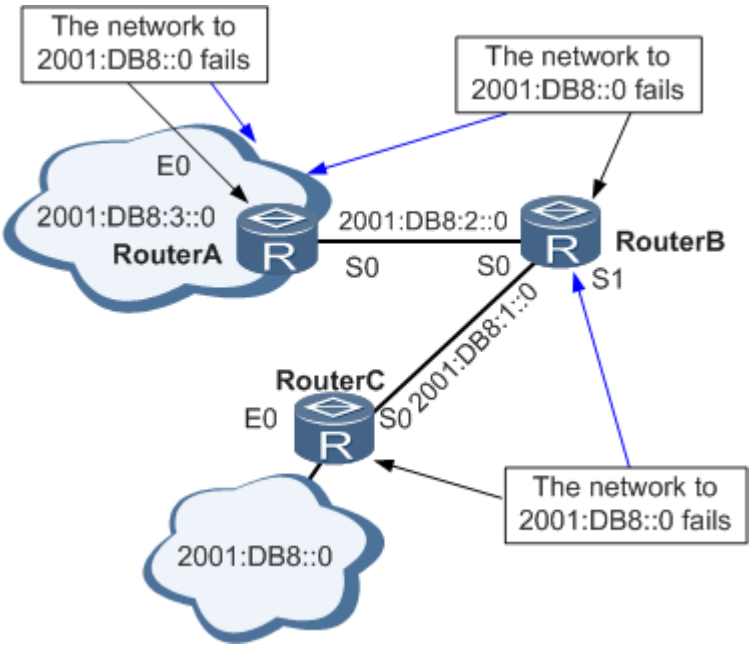
11.12.4.6 Triggered Update

Triggered update allows a device to advertise the routing information changes immediately, which speeds up network convergence.

Triggered update shortens the network convergence time. When the local routing information changes, the local router immediately notifies its neighbors of the changes in routing information rather than wait for a periodic update.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-85 Networking for triggered update



As shown in **Figure 11-85**, when network 2001:DB8::0 is unreachable, Router C learns the information first. Usually, the route update message is periodically sent to neighbor routers. For example, RIPng sends the route update message every 30s. If Router C receives an Update packet from Router B within 30s while Router C is still waiting to send Update packets, Router C learns the incorrect route to 2001:DB8::0 from Router B. In this case, the next hops of the routes from Routers B and C to 2001:DB8::0 point to each other and form a loop. If Router C sends an Update packet to Router B immediately after it detects a network fault, Router B can rapidly update its routing table, which prevents routing loops.

In addition, if the next hop of a route becomes unavailable due to a link failure, the local Router sets the metric of the route to 16 and then advertises the route immediately to its neighbors. This process is called route poisoning.

11.12.4.7 Route Summarization

Background

On large networks, the Routing Information Protocol next generation (RIPng) routing table of each device contains a large number of routes, which consumes lots of system resources. In addition, if a specific link connected to a device within an IP address range frequently alternates between Up and Down, route flapping occurs.

To address these problems, RIPng route summarization was introduced. With RIPng route summarization, a device summarizes routes destined for different subnets of a network segment into one route destined for one network segment and then advertises the summarized route to other network segments. RIPng route summarization reduces the number of routes in the routing table, minimizes system resource consumption, and prevents route flapping.

 Copyright © Huawei Technologies Co., Ltd.

Implementation

RIPng route summarization is interface-based. After RIPng route summarization is enabled on an interface, the interface summarizes routes based on the longest matching rule and then advertises the summarized route. The smallest metric among the specific routes for the summarization is used as the metric of the summarized route.

For example, an interface has two routes: 2001:db8:11::24 with metric 2 and 2001:db8:12::34 with metric 3. After RIPng route summarization is enabled on the interface, the interface summarizes the two routes into the route 2001:db8::0/32 with metric 2 and then advertises it.

11.12.4.8 Multi-Process and Multi-Instance

RIPng supports multi-process and multi-instance, which simplifies network management and improves service control efficiency. Multi-process allows a set of interfaces to be associated with a specific RIPng process, which ensures that the specific RIPng process performs all the protocol operations only on this set of interfaces. Therefore, multiple RIPng processes can run on one router, and each process manages a unique set of interfaces. In addition, the routing data of each RIPng process is independent; however, processes can import routes from each other.

On routers that support VPN, each RIPng process is associated with a specific VPN instance. Therefore, all the interfaces associated with the RIPng process need to be associated with the RIPng process-related VPN instance.

11.12.4.9 Hot Standby

Devices of a distributed architecture support RIPng Hot Standby (HSB). In the RIPng HSB process, RIPng backs up RIPng configuration from the Active Main Board (AMB) to the Standby Main Board (SMB). Whenever the AMB fails, the SMB becomes active to take over the work. RIPng re-sends synchronization requests to ensure that RIPng works properly.

11.12.4.10 IPSec Authentication

Background

As networks develop, network security has become an increasing concern. Internet Protocol Security (IPsec) authentication is used to authenticate Routing Information Protocol next generation (RIPng) packets for the network security concern. The packets that fail to be authenticated are discarded, which prevents data transmitted based on TCP/IP from being intercepted, tampered with, or attacked.

Implementation

IPSec has an open standard architecture and ensures secure packet transmission on the Internet by encrypting packets. RIPng IPSec provides a complete set of security protection mechanisms to authenticate RIPng packets, which prevents devices from being attacked by forged RIPng packets.

 Copyright © Huawei Technologies Co., Ltd.

IPSec includes a set of protocols that are used at the network layer to ensure data security, such as Authentication Header (AH), Encapsulating Security Payload (ESP), and Internet Key Exchange (IKE). The three protocols are described as follows:

- AH: A protocol that provides data origin authentication, data integrity check, and anti-replay protection. AH does not encrypt packets to be protected.
- ESP: A protocol that provides IP packet encryption and authentication mechanisms besides the functions provided by AH. The encryption and authentication mechanisms can be used together or independently.

  NOTE

  AH and ESP can be used together or independently.

Benefits

RIPng IPSec offers the following benefits:

- Improves carriers' reputation and competitiveness by preventing services from being tampered with or attacked by unauthorized users.
- Ensures confidentiality and integrity of user packets.

11.12.4.11 RIPng NSR

Background

As networks develop, the demand for data, audio, and video services is growing, which impose increasing requirements on IP network reliability. If an AMB/SMB switchover is performed on a device during a maintenance operation or a single point of failure occurs, the device may fail to converge routing information, which may result in traffic loss or even a network breakdown. Non-stop routing (NSR) can address this problem and ensure uninterrupted forwarding of key services.

Related Concepts

- High availability (HA): supports data backup between the AMB and SMB.
- Non-stop forwarding (NSF): enables a node to use the GR mechanism to ensure uninterrupted transmission during an AMB/SMB switchover.
- NSR: allows a standby control plane to take over traffic from an active control plane if the active control plane fails, preventing the control planes of neighbors from detecting the fault.
- AMB and SMB: run control plane processes.

Implementation

With Routing Information Protocol next generation (RIPng) NSR, RIPng real-time data is synchronized between the AMB and SMB. After an AMB/SMB switchover is performed on a device, the SMB takes over services from the AMB, and neighbors are unaware of the local fault. After the switchover, the new AMB restores IS-IS immediately based on the synchronized IS-IS real-time data. Therefore, neighbors are unaware of the switchover as well. RIPng NSR requires synchronization of the following data:

 Copyright © Huawei Technologies Co., Ltd.

- All configuration data, such as information about neighbors, timer parameters, and process configurations.
- Dynamic data, such as the interface parameters and state, and information about neighbors and the link state database (LSDB).

Usage Scenario

NSR minimizes the impact of control plane faults and prevents route flapping on networks that require high reliability.

Benefits

NSR improves network reliability and ensures uninterrupted traffic forwarding.

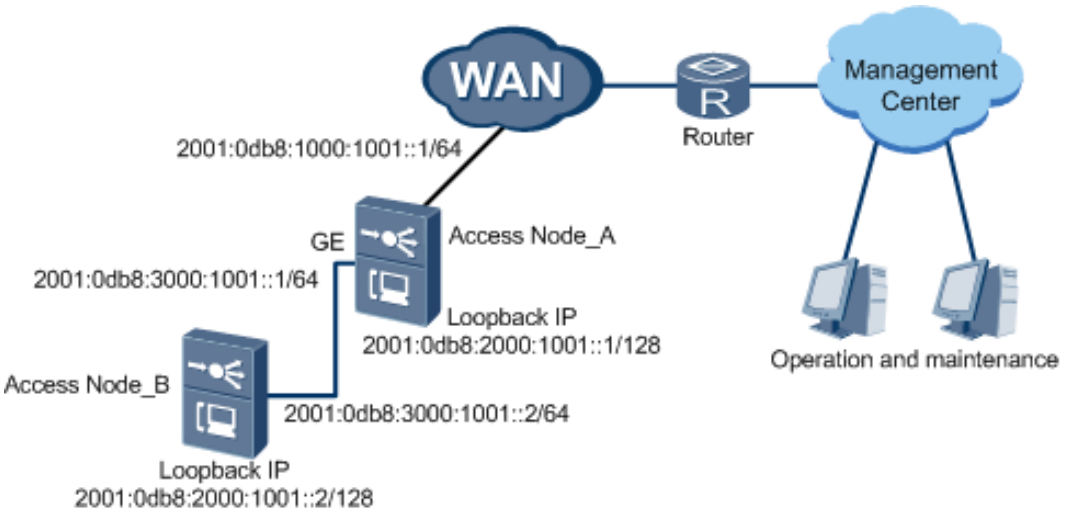
11.12.4.12 Configuration Example of RIPng

This topic provides an example for configuring RIPng on the MA5800.

Service Requirements

- Access Node_A is cascaded with Access node_B through port 0/9/1, and uses port 0/9/0 to transmit services in the upstream. Besides, it connects to the management center network through the WAN.
- RIPng is enabled on Access Node_A and Access Node_B so that the administrator can access Access Node_A and Access Node_B through the RIPng route. Then, you can operate and maintain Access Node_A and Access Node_B.

Figure 11-86 Example network for configuring RIPng



Data Plan

Table 11-21 provides the data plan for configuring RIPng.

 Copyright © Huawei Technologies Co., Ltd.

Table 11-21 Data plan for configuring RIPng

Item	Data
Access node_A	Upstream port: 0/9/0 Administration VLAN: smart VLAN 100 IPv6 address of the Layer 3 interface in the administration VLAN: 2001:0db8:1000:1001::1/64 Loopback interface address: 2001:0db8:2000:1001::1/128
	RIPng route filtering policy: filtering routes based on the IPv6 address prefix list "abc". Only the routes with the IP addresses 2001:0db8:2000:1001::1 and 2001:0db8:2000:1001::2 can be advertised through the Layer 3 interface of VLAN 100.
	Cascading port: 0/9/0 Cascading administration VLAN: smart VLAN 10 IPv6 address of the Layer 3 interface in the cascading administration VLAN: 2001:0db8:3000:1001::1/64
Access node_B	Cascading port: 0/9/1 Administration VLAN: smart VLAN 10 IPv6 address of the Layer 3 interface in the administration VLAN: 2001:0db8:3000:1001::2/64 Loopback interface address: 2001:0db8:2000:1001::2/128
	RIPng route filtering policy: filtering routes based on the IPv6 address prefix list "abc". Only the route with the IP address 2001:0db8:2000:1001::2 can be advertised through the Layer 3 interface of VLAN 10.

Procedure

- Configure Access Node_A.
  a. Configure the RIPng-supported Layer 3 interface.

```
huawei(config)#vlan 100 smart
huawei(config)#port vlan 100 0/9 0
huawei(config)#interface vlanif 100
huawei(config-if-vlanif100)#ipv6 enable
huawei(config-if-vlanif100)#ipv6 address 2001:0db8:1000:1001::1 64
huawei(config-if-vlanif100)#quit
huawei(config)#interface loopBack 0
huawei(config-if-loopback0)#ipv6 enable
huawei(config-if-loopback0)#ipv6 address 2001:0db8:2000:1001::1/128
huawei(config-if-loopback0)#quit
```

  b. Enable RIPng.

```
huawei(config)#ripng 1
huawei(config-ripng-1)#quit
huawei(config)#interface vlanif 100
huawei(config-if-vlanif100)#ripng 1 enable
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-if-vlanif100)#quit
huawei(config)#interface loopBack 0
huawei(config-if-loopback0)#ripng 1 enable
huawei(config-if-loopback0)#quit
```

c. Configure the route filtering policy.

```
huawei(config)#ip ipv6-prefix abc permit 2001:0db8:2000:1001::1 128
huawei(config)#ip ipv6-prefix abc permit 2001:0db8:2000:1001::2 128
huawei(config)#ripng 1
huawei(config-ripng-1)#filter-policy ipv6-prefix abc export
huawei(config-ripng-1)#quit
```

d. Configure the cascading port.

```
huawei(config)#vlan 10 smart
huawei(config)#port vlan 10 0/9 1
huawei(config)#interface mpu 0/9
huawei(config-if-mpu-0/9)#network-role 1 cascade
huawei(config-if-mpu-0/9)#quit
huawei(config)#interface vlanif 10
huawei(config-if-vlanif10)#ipv6 enable
huawei(config-if-vlanif10)#ipv6 address 2001:0db8:3000:1001::1 64
huawei(config-if-vlanif10)#quit
```

e. Enable RIPng on the cascading port.

```
huawei(config)#interface vlanif 10
huawei(config-if-vlanif10)#ripng 1 enable
huawei(config-if-vlanif10)#quit
```

f. Save the data.

```
huawei(config)#save
```

- Configure Access Node_B.

a. Configure the RIPng-supported Layer 3 interface.

```
huawei(config)#vlan 10 smart
huawei(config)#port vlan 10 0/9 0
huawei(config)#interface vlanif 10
huawei(config-if-vlanif10)#ipv6 enable
huawei(config-if-vlanif10)#ipv6 address 2001:0db8:3000:1001::2 64
huawei(config-if-vlanif10)#quit
huawei(config)#interface loopback 0
huawei(config-if-loopback0)#ipv6 enable
huawei(config-if-loopback0)#ipv6 address 2001:0db8:2000:1001::2 128
huawei(config-if-loopback0)#quit
```

b. Enable RIPng.

```
huawei(config)#ripng 1
huawei(config-ripng-1)#quit
huawei(config)#interface vlanif 10
huawei(config-if-vlanif10)#ripng 1 enable
huawei(config-if-vlanif10)#quit
huawei(config)#interface loopback 0
huawei(config-if-loopback0)#ripng 1 enable
huawei(config-if-loopback0)#quit
```

c. Configure the route filtering policy.

```
huawei(config)#ip ipv6-prefix abc permit 2001:0db8:2000:1001::2 128
huawei(config)#ripng 1
huawei(config-ripng-1)#filter-policy ipv6-prefix abc export
huawei(config-ripng-1)#quit
```

d. Save the data.

```
huawei(config)#save
```

----End

 Copyright © Huawei Technologies Co., Ltd.

Result

The maintenance terminal of the administration center can access Access Node_A and Access Node_B, and operate and maintain the two devices.

Configuration File

Configuration on Access Node_A

```
vlan 100 smart
port vlan 100 0/9 0
interface vlanif 100
ipv6 enable
ipv6 address 2001:0db8:1000:1001::1 64
quit
interface loopBack 0
ipv6 enable
ipv6 address 2001:0db8:2000:1001::1/128
quit
ripng 1
quit
interface vlanif 100
ripng 1 enable
quit
interface loopBack 0
ripng 1 enable
quit
ip ipv6-prefix abc permit 2001:0db8:2000:1001::1 128
ip ipv6-prefix abc permit 2001:0db8:2000:1001::2 128
ripng 1
filter-policy ipv6-prefix abc export
quit
vlan 10 smart
port vlan 10 0/9 1
interface mpu 0/9
network-role 1 cascade
quit
interface vlanif 10
ipv6 enable
ipv6 address 2001:0db8:3000:1001::1 64
ripng 1 enable
quit
save
```

Configuration on Access node_B

```
vlan 10 smart
port vlan 10 0/9 0
interface vlanif 10
interface vlanif 10
ipv6 enable
ipv6 address 2001:0db8:3000:1001::2 64
quit
interface loopback 0
ipv6 enable
ipv6 address 2001:0db8:2000:1001::2 128
quit
ripng 1
quit
interface vlanif 10
ripng 1 enable
quit
interface loopback 0
ripng 1 enable
quit
ip ipv6-prefix abc permit 2001:0db8:2000:1001::2 128
ripng 1
filter-policy ipv6-prefix abc export
```

 Copyright © Huawei Technologies Co., Ltd.

```
quit
save
```

11.12.4.13 RIPng Reference Standards and Protocols

The following table lists the references.

Document No.	Document Name	Remarks
RFC 2080	RIPng for IPv6	-

Document No.	Document Name	Protocol Compliance
RFC 2080	RIPng for IPv6	Fully compliant.

11.12.5 IS-IS

11.12.5.1 Introduction

Definition

Intermediate System to Intermediate System (IS-IS) is a dynamic routing protocol initially designed by the International Organization for Standardization (ISO) for its Connectionless Network Protocol (CLNP).

To support IP routing, the Internet Engineering Task Force (IETF) extends and modifies IS-IS in RFC 1195, which enables IS-IS to be applied to both TCP/IP and Open System Interconnection (OSI) environments. This type of IS-IS is called Integrated IS-IS or Dual IS-IS.

In this document, IS-IS refers to Integrated IS-IS, unless otherwise stated.

Purpose

As an Interior Gateway Protocol (IGP), IS-IS is used in Autonomous Systems (ASs). IS-IS is a link state protocol, and it uses the Shortest Path First (SPF) algorithm to calculate routes.

11.12.5.2 Principles

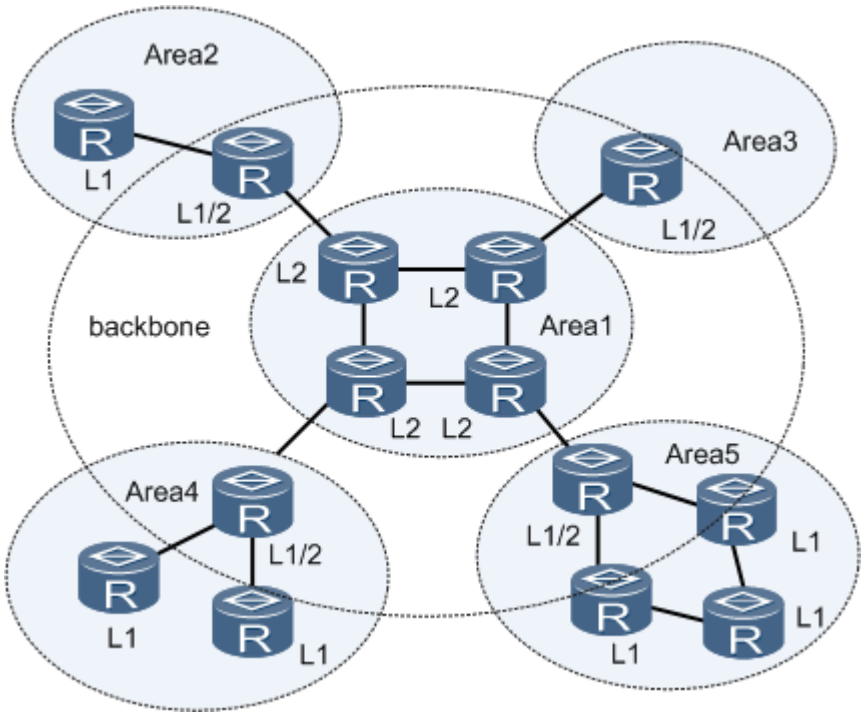
 Copyright © Huawei Technologies Co., Ltd.

11.12.5.2.1 Basic Concepts of IS-IS

IS-IS Areas

To support large-scale routing networks, IS-IS adopts a two-level structure in a routing domain. A large domain is divided into areas. **Figure 11-87** shows an IS-IS network. The entire backbone area covers all Level-2 routers in area 1 and Level-1-2 routers in other areas. Three types of routers on the IS-IS network are described as follows:

Figure 11-87 IS-IS topology



- Level-1 router

  A Level-1 device manages intra-area routing. It establishes neighbor relationships with only the Level-1 and Level-1-2 devices in the same area and maintains a Level-1 LSDB. The LSDB contains routing information in the local area. A packet to a destination beyond this area is forwarded to the nearest Level-1-2 device.

- Level-2 router

  A Level-2 device manages inter-area routing. It can establish neighbor relationships with all Level-2 devices and Level-1-2 devices, and maintains a Level-2 LSDB which contains inter-area routing information.

  All Level-2 devices form the backbone network of the routing domain. Level-2 neighbor relationships are set up between them. They are responsible for communications between areas. The Level-2 devices in the routing domain must be in succession to ensure the continuity of the backbone network. Only Level-2 devices can exchange data packets or routing information with the devices beyond the routing domain.

- Level-1-2 router

 Copyright © Huawei Technologies Co., Ltd.

A device, which can establish neighbor relationships with both Level-1 devices and Level-2 devices, is called a Level-1-2 device. A Level-1-2 device can establish Level-1 neighbor relationships with Level-1 devices and Level-1-2 devices in the same area. It can also establish Level-2 neighbor relationships with Level-2 devices and Level-1-2 devices in other areas. Level-1 devices can be connected to other areas only through Level-1-2 devices.

A Level-1-2 device maintains two LSDBs: a Level-1 LSDB and a Level-2 LSDB. The Level-1 LSDB is used for intra-area routing, while the Level-2 LSDB is used for inter-area routing.

NOTE

Level-1 routers in different areas cannot establish neighbor relationships. Level-2 routers can establish neighbor relationships with each other, regardless of the areas to which the Level-2 routers belong.

In general, Level-1 devices are located within an area, Level-2 devices are located between areas, and Level-1-2 devices are located between Level-1 devices and Level-2 devices.

Interface level

A Level-1-2 device may need to establish only a Level-1 adjacency with a neighbor and establish only a Level-2 adjacency with another neighbor. In this case, you can set the level of an interface to control the setting of adjacencies on the interface. Specifically, only Level-1 adjacencies can be established on a Level-1 interface, and only Level-2 adjacencies can be established on a Level-2 interface.

Address Structure of IS-IS

In OSI, the NSAP is used to locate resources. The ISO adopts the address structure shown in **Figure 11-88**. An NSAP is composed of the Initial Domain Part (IDP) and the Domain Specific Part (DSP). IDP is the counterpart of network ID in an IP address, and DSP is the counterpart of the subnet number and host address in an IP address.

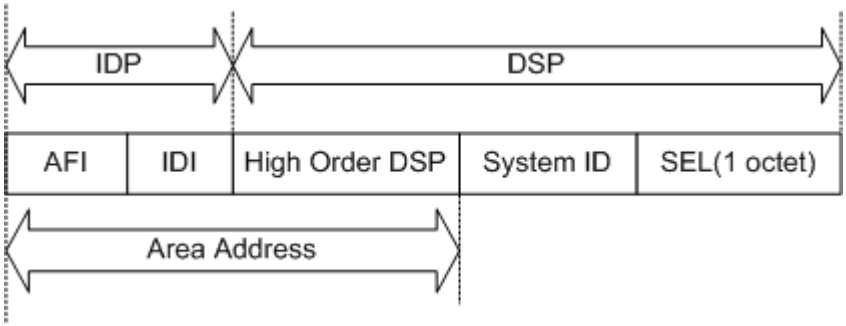
As defined by the ISO, the IDP consists of the Authority and Format Identifier (AFI) and Initial Domain Identifier (IDI). AFI specifies the address assignment mechanism and the address format; the IDI identifies a domain.

The DSP consists of the High Order DSP (HODSP), system ID, and NSAP Selector (SEL). The HODSP is used to divide areas; the system ID identifies a host; the SEL indicates the service type.

The lengths of the IDP and DSP are variable. The length of the NSAP varies from 8 bytes to 20 bytes.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-88 IS-IS address structure



- Area address

  An IDP and HODSP of the DSP can identify a routing domain and the areas in a routing domain; therefore, the combination of the IDP and HODSP is referred to as an area address, equal to an area ID in OSPF. An area address is used to uniquely identify an area in a routing domain. The area addresses of routers in the same Level-1 area must be the same, while the area addresses of routers in the Level-2 area can be different.

  In general, a router can be configured with only one area address. The area address of all nodes in an area must be the same. In the implementation of a device, an IS-IS process can be configured with a maximum of three area addresses to support seamless combination, division, and transformation of areas.

- System ID

  A system ID uniquely identifies a host or a router in an area. In the device, the length of the system ID is 48 bits (6 bytes).

  In actual applications, a router ID corresponds to a system ID. If a device uses the IP address of Loopback 0 (192.168.1.1) as its router ID, its system ID used in IS-IS can be obtained in the following steps:

  – Extend each part of the IP address 192.168.1.1 to 3 digits and add 0 or 0s to the front of the part that is shorter than 3 digits.
  – Divide the extended address 192.168.001.001 into three parts, with each part consisting of 4 decimal digits.
  – The reconstructed 1921.6800.1001 is the system ID.

  There are many ways to specify a system ID. You need to ensure that the system ID uniquely identifies a host or a device.

- SEL

  The role of an SEL (also referred to as NSAP Selector or N-SEL) is similar to that of the "protocol identifier" of IP. A transport protocol matches an SEL. The SEL is "00" in IP.

- NET

  A Network Entity Title (NET) indicates the network layer information of an IS itself. It does not contain the transport layer information (SEL = 0). A NET can be regarded as a special NSAP. The length of the NET field is the same as that of an NSAP, varying from 8 bytes to 20 bytes. For example, in NET ab.cdef.1234.5678.9abc.00, the area is ab.cdef, the system ID is 1234.5678.9abc, and the SEL is 00.

 Copyright © Huawei Technologies Co., Ltd.

In general, an IS-IS process is configured with only one NET. When areas need to be redefined, for example, areas need to be combined or an area needs to be divided into sub-areas, you can configure multiple NETs.

NOTE

A maximum of three area addresses can be configured in an IS-IS process, and therefore, you can configure only a maximum of three NETs. When you configure multiple NETs, ensure that their system IDs are the same.

The routers in an area must have the same area address.

IS-IS Network Types

IS-IS supports the following types of networks:

- Broadcast network
- Point-to-point (P2P) network

11.12.5.2.2 Basic Protocols of IS-IS

Related Concepts

DIS and Pseudo Node

A Designated Intermediate System (DIS) is an intermediate router elected in IS-IS communication. A pseudo node simulates a virtual node on a broadcast network and is not a real router. In IS-IS, a pseudo node is identified by the system ID and 1-byte circuit ID (a non-zero value) of a DIS.

The DIS is used to create and update pseudo nodes and generate the **link state protocol data units (LSPs)** of pseudo nodes. The routers advertise a single link to a pseudo node and obtain routing information about the entire network through the pseudo node. The router does not need to exchange packets with all the other routers on the network. Using the DIS and pseudo nodes simplifies network topology and reduces the length of LSPs generated by routers. When the network changes, fewer LSPs are generated. As a result, the SPF consumes fewer resources.

SPF Algorithm

The SPF algorithm, also named Dijkstra's algorithm, is used in a link-state routing protocol to calculate the shortest paths to other nodes on a network. In the SPF algorithm, a local router takes itself as the root and generates a shortest path tree (SPT) based on the network topology to calculate the shortest path to every destination node on a network. In IS-IS, the SPF algorithm runs separately in Level-1 and Level-2 databases.

Implementation

All routers on the IS-IS network communicate through the following steps:

- **Establishment of IS-IS Neighbor Relationships**
- **LSDB Synchronization**
- **Route Calculation**

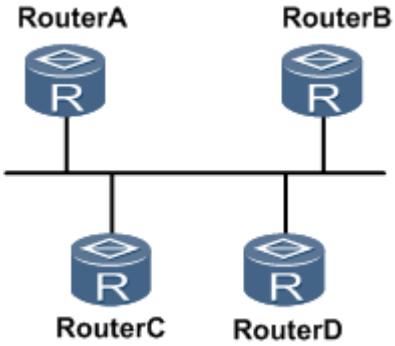
Establishment of IS-IS Neighbor Relationships

 Copyright © Huawei Technologies Co., Ltd.

On different types of networks, the modes for establishing IS-IS neighbor relationships are different.

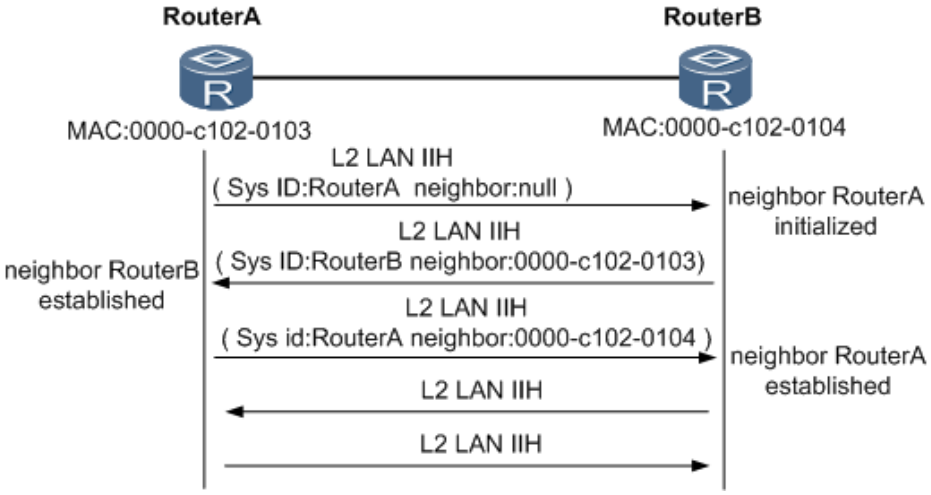
- Establishment of a neighbor relationship on a broadcast link

Figure 11-89 Networking for a broadcast link



Router A, Router B, Router C, and Router D are Level-2 routers. Router A is newly added to the broadcast network. **Figure 11-90** demonstrates the process of establishing the neighbor relationship between Router A and Router B, the process of establishing the neighbor relationship between Router A and Router C or Router D is similar to that between Router A and Router B.

Figure 11-90 Establishing a neighbor relationship on a broadcast link



As shown in **Figure 11-90**, the process for establishing a neighbor relationship on a broadcast link consists of the following phases:

- Router A broadcasts a Level-2 local area network (LAN) **IS-to-IS Hello PDU (IIH)**. After Router B receives the IIH, Router B detects that the neighbor field in the IIH does not contain its media access control (MAC) address, and sets its neighbor status with Router A to **Initial**.
- Router B replies a Level-2 LAN IIH to Router A. After Router A receives the IIH, Router A detects that the neighbor field in the IIH contains its MAC address, and sets its neighbor status with Router B to **Up**.

 Copyright © Huawei Technologies Co., Ltd.

– Router A sends a Level-2 LAN IIH to Router B. After Router B receives the IIH, Router B detects that the neighbor field in the IIH contains its MAC address, and sets its neighbor status with Router A to **Up**.

DIS Election

On a broadcast network, any two routers exchange information. If n routers are available on the network, n x (n - 1)/2 adjacencies must be established. Each status change of a router is transmitted to other routers, which wastes bandwidth resources. IS-IS resolves this problem by introducing the DIS. All routers send information to the DIS, which then broadcasts the network link status. Using the DIS and pseudo nodes simplifies network topology and reduces the length of LSPs generated by routers. When the network changes, fewer LSPs are generated. As a result, the SPF consumes fewer resources.

A DIS is elected after a neighbor relationship is established. Level-1 and Level-2 DISs are elected separately. You can configure different priorities for DISs at different levels. In DIS election, a Level-1 priority and a Level-2 priority are specified for every interface on every router. A router uses every interface to send IIHs and advertises its priorities in the IIHs to neighboring routers. The higher the priority, the higher the probability of being elected as the DIS. If there are multiple routers with the same highest priority on a broadcast network, the one with the largest MAC address is elected. The DISs at different levels can be the same router or different routers.

In the DIS election procedure, IS-IS is different from Open Shortest Path First (OSPF) in the following aspects:

– The router with the priority of 0 also takes part in the DIS election.

– When a new router that meets the requirements of being a DIS is added to the broadcast network, the router is selected as the new DIS, which triggers a new round of LSP flooding.

- Establishment of a neighbor relationship on a P2P link

  The establishment of a neighbor relationship on a P2P link is different from that on a broadcast link. A neighbor relationship on a P2P link can be established in 2-way or 3-way mode, as shown in **Table 11-22**. By default, the 3-way handshake mechanism is used to establish a neighbor relationship on a P2P link.

 Copyright © Huawei Technologies Co., Ltd.

Table 11-22 Comparison between 2-way mode and 3-way mode

Mode	Description	Advantages and Disadvantages	Reliability
2-way mode	When a router receives an IIH, it unidirectionally sets up a neighbor relationship.	Disadvantages: • The unstable link status causes the loss of **complete sequence numbers protocol data units (CSNPs)** that are sent once an adjacency is set up. As a result, the link state databases (LSDBs) of two neighboring routers are not synchronized during the LSP update period. • If two or more links exist between two routers, an adjacency can still be set up when one link is Down and another is Up in the same direction. A router that fails to detect the faulty link may also forward packets over this link.	Low

 Copyright © Huawei Technologies Co., Ltd.

Mode	Description	Advantages and Disadvantages	Reliability
3-way mode	A neighbor relationship is established after IIHs are sent three times.	Advantages: A neighbor relationship is established only when both ends are Up. This mechanism ensures that packets are transmitted securely.	High

IS-IS Strict Neighbor Check

In IS-IS, IPv4 and IPv6 share a standard topology. When both IPv4 and IPv6 are deployed in a standard topology, a problem can occur: if only IPv4 goes Up, IPv6 traffic is interrupted; if only IPv6 goes Up, IPv4 traffic is interrupted. Another problem occurs when a primary link recovers after a switchover occurs between primary and backup links. In this scenario, IPv4 traffic is switched to the primary link first because IPv4 goes Up faster than IPv6. IPv6 is also switched but discarded because IPv6 links are not Up. IS-IS strict neighbor check can resolve these problems.

IS-IS must check the following items when establishing a neighbor relationship:

- Whether the address family configured on the local router is the same as that carried in a received IIH. The address family status of neighboring routers is set to Up only when both are the same.
- Whether the address family configured in the standard topology is Up in the case of establishing a standard topology neighbor relationship identified by MT0, which is carried in an IIH. The address family configured in the standard topology goes Up only when all address families go Up.

LSDB Synchronization

IS-IS is a link-state protocol. An IS-IS router obtains first-hand information from other routers running link-state protocols. Every router generates information about itself, directly connected networks, and links between itself and directly connected networks. The router then sends the generated information to other routers through adjacent routers. Every router saves link state information without modifying it. Finally, every router has the same network interworking information, and LSDB synchronization is complete. The process of synchronizing LSDBs is called LSP flooding. In LSP flooding, a router sends an LSP to its neighbors and the neighbors send the received LSP to their neighbors except the router that first sends the LSP. The LSP is flooded among the routers at the same level. This implementation allows each router at the same level to have the same LSP information and keep a synchronized LSDB.

All routers in the IS-IS routing domain can generate LSPs. The following events trigger the generation of a new LSP:

 Copyright © Huawei Technologies Co., Ltd.

- Neighbor goes Up or Down.
- related interface goes Up or Down.
- Imported IP routes change.
- Inter-area IP routes change.
- A new metric value is configured for an interface.
- Periodic updates occur.

A router processes a received LSP as follows:

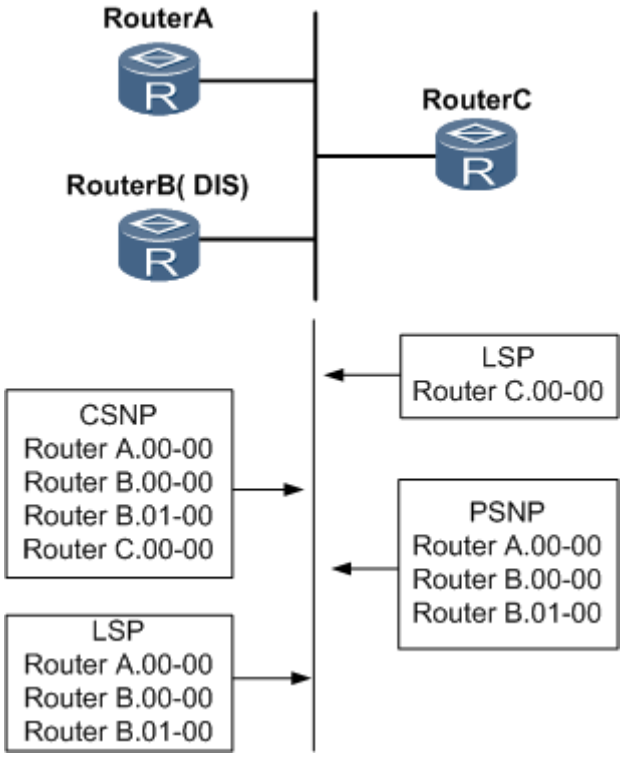
- Updating the LSDB on a broadcast link

  The DIS updates the LSDB to synchronize LSDBs on a broadcast network. **Figure 11-91** shows the process of synchronizing LSDBs on a broadcast network.

  a. When the DIS receives an LSP, it searches the LSDB for the related records. If the DIS does not find the LSP in its LSDB, it adds the LSP to its LSDB and broadcasts the new LSDB.
  b. If the sequence number of the received LSP is greater than that of the local LSP, the DIS replaces the local LSP with the received LSP in the LSDB and broadcasts the new LSDB.
  c. If the sequence number of the received LSP is less than that of the local LSP, the DIS sends the local LSP in the LSDB to the inbound interface.
  d. If the sequence number of the received LSP is equal to that of the local LSP, the DIS compares the Remaining Lifetime of the two LSPs. If Remaining Lifetime of the received LSP is 0, the DIS replaces the LSP with the received LSP, and broadcasts the new LSDB. If the Remaining Lifetime of local LSP is 0, the DIS sends the LSP to the inbound interface.
  e. If the sequence number of the received LSP and the local LSP in the LSDB are the same and neither Remaining Lifetime is 0, the DIS compares the checksum of the two LSPs. If the received LSP has a greater checksum than that of the local LSP in the LSDB, the DIS replaces the local LSP in the LSDB with the received LSP and advertises the new LSDB. If the received LSP has a smaller checksum than that of the local LSP in the LSDB, the DIS sends the local LSP in the LSDB to the inbound interface.
  f. If the checksums of the received LSP and the local LSP are the same, the LSP is not forwarded.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-91 Process of updating the LSDB on a broadcast link



- Updating the LSDB on a P2P link
  a. If the sequence number of the received LSP is greater than that of the local LSP in the LSDB, the router adds the received LSP to its LSDB. The router then sends a **PSNP packet** to acknowledge the received LSP and sends the LSP to all its neighbors except the neighbor that sends the LSP.
  b. If the sequence number of the received LSP is less than that of the local LSP, the router directly sends its LSP to the neighbor and waits for a PSNP from the neighbor as an acknowledgement.
  c. If the sequence number of the received LSP is the same as that of the local LSP in the LSDB, the router compares the Remaining Lifetimes of the two LSPs. If Remaining Lifetime of the received LSP is 0, the router adds the LSP to its LSDB. The router then sends a PSNP to acknowledge the received LSP. If Remaining Lifetime of the local LSP is 0, the router directly sends the local LSP to the neighbor and waits for a PSNP from the neighbor.
  d. If the sequence number of the received LSP and the local LSP in the LSDB are the same, and neither Remaining Lifetime is 0, the router compares the checksum of the two LSPs. If the received LSP has a greater checksum than that of the local LSP, the router adds the received LSP to its LSDB. The router then sends a PSNP to acknowledge the received LSP. If the received LSP has a smaller checksum than that of the local LSP, the router directly sends the local LSP to the neighbor and waits for a PSNP from the neighbor. At last, the router sends the LSP to all its neighbors except the neighbor that sends the LSP.
  e. If the checksums of the received LSP and the local LSP are the same, the LSP is not forwarded.

Route Calculation

 Copyright © Huawei Technologies Co., Ltd.

When LSDB synchronization is complete and network convergence is implemented, IS-IS performs SPF calculation by using LSDB information to obtain the SPT. IS-IS uses the SPT to create a forwarding database (a routing table).

In IS-IS, link costs are used to calculate shortest paths. The default cost for an interface on a Huawei router is 10. The cost is configurable. The cost of a route is the sum of the cost of every outbound interface along the route. There may be multiple routes to a destination, among which the route with the smallest cost is the optimal route.

Level-1 routers can also calculate the shortest path to Level-2 routers to implement inter-area route selection. When a Level-1-2 router is connected to other areas, the router sets the value of the attachment (ATT) bit in its LSP to 1 and send the LSP to neighboring routers. In the route calculation process, a Level-1 router selects the nearest Level-1-2 router as an intermediate router between the Level-1 and Level-2 areas.

11.12.5.2.3 IS-IS Routing Information Control

IS-IS routes calculated using the shortest path first (SPF) algorithm may bring about some problems. For example, too many routing entries slow down route lookup, or link usage is unbalanced. As a result, IS-IS routing cannot meet carriers' network planning and traffic management requirements.

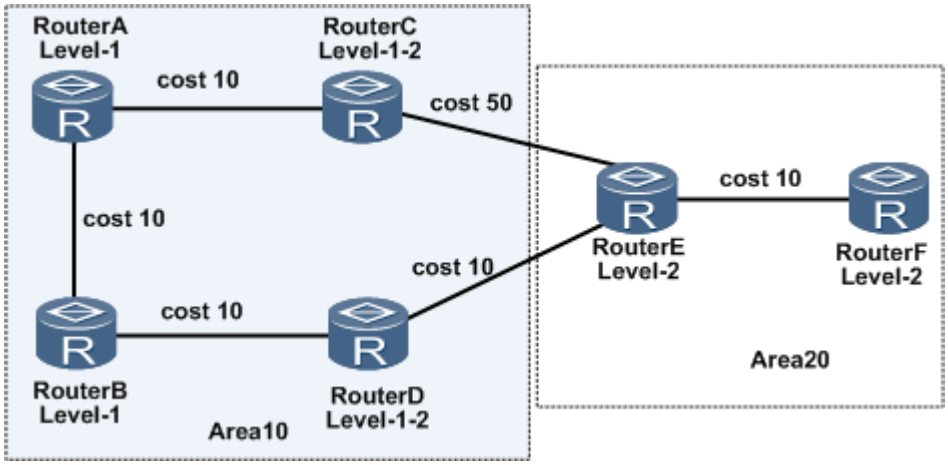
IS-IS routing information control can refine control over route selection using the following methods.

Route Leaking

When Level-1 and Level-2 areas both exist on an IS-IS network, Level-2 routers do not advertise the learned routing information about a Level-1 area and the backbone area to any other Level-1 area by default. Therefore, Level-1 routers do not know the routing information beyond the local area. As a result, the Level-1 routers cannot select the optimal routes to the destination beyond the local area.

With route leaking, Level-1-2 routers can select routes using ACLs, routing policies, or tags and advertise the selected routes of other Level-1 areas and the backbone area to the Level-1 area. **Figure 11-92** shows the typical networking for route leaking.

Figure 11-92 Typical networking for route leaking



 Copyright © Huawei Technologies Co., Ltd.

- Router A, Router B, Router C, and Router D belong to area 10. Router A and Router B are Level-1 routers. Router C and Router D are Level-1-2 routers.
- Router E and Router F belong to area 20 and are Level-2 routers.

If Router A sends a packet to Router F, the selected optimal route should be Router A -> Router B -> Router D -> Router E -> Router F because its cost is 40 (10 + 10 + 10 + 10 = 40) which is less than that of Router A -> Router C -> Router E -> Router F (10 + 50 + 10 = 70). However, if you check routes on Router A, you can find that the selected route is Router A -> Router C -> Router E -> Router F, which is not the optimal route from Router A to Router F.

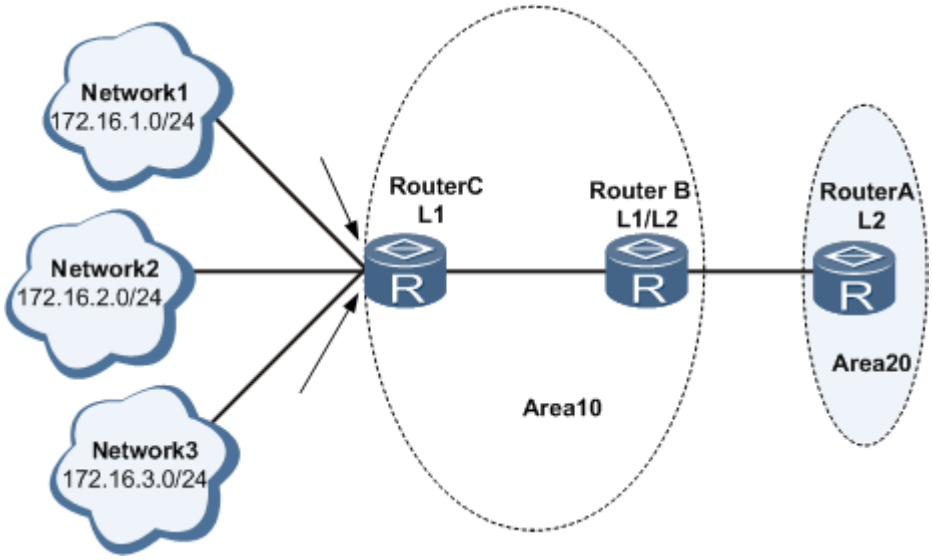
This is because Router A does not know the routes beyond the local area, and therefore, the packets sent by Router A to other network segments are sent through the default route generated by the nearest Level-1-2 device.

In this case, you can enable route leaking on the Level-1-2 devices (Router C and Router D). Then, check the route and you can find that the selected route is Router A -> Router B -> Router D -> Router E -> Router F.

Route Summarization

On a large-scale IS-IS network, links connected to devices within an IP address range may alternate between Up and Down. With route summarization, multiple routes with the same IP prefix are summarized into one route, which prevents route flapping, reduces routing entries and system resource consumption, and facilitates route management. **Figure 11-93** shows the typical networking for route summarization.

Figure 11-93 Typical networking for route summarization



- Router A, Router B, and Router C use IS-IS to communicate with each other.
- Router A belongs to area 20, and Router B and Router C belong to area 10.
- Router A is a Level-2 router. Router B is a Level-1-2 router. Router C is a Level-1 router.

 Copyright © Huawei Technologies Co., Ltd.

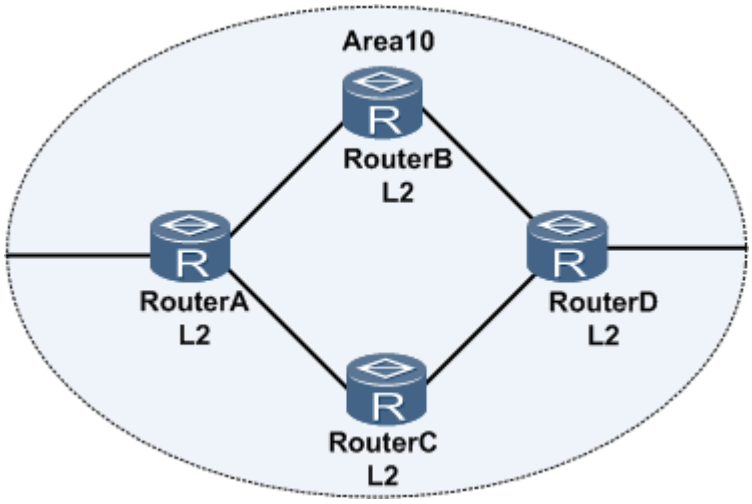
- Router B maintains Level-1 and Level-2 LSDBs and leaks the routes to three network segments (172.16.1.0/24, 172.16.2.0/24, and 172.16.3.0/24) from the Level-1 area to the Level-2 area. If a link fault causes the Router C interface with IP address 172.16.1.1/24 to frequently alternate between Up and Down, the status change is advertised to the Level-2 area, triggering frequent LSP flooding and SPF calculation on Router A. As a result, the CPU usage on Router A increases, and even network flapping occurs.

  On Router B, routes to the three network segments in the Level-1 area are summarized to one route to 172.16.0.0/16, which reduces the number of routing entries on Router B and minimizes the impact of route flapping in the Level-1 area on route convergence in the Level-2 area.

Load Balancing

If multiple equal-cost routes are available on a network, they can load-balance traffic, which improves link usage and prevents network congestion caused by link overload. **Figure 11-94** shows the typical networking for load balancing.

Figure 11-94 Typical networking for load balancing



- Router A, Router B, Router C, and Router D communicate with each other on an IP network using IS-IS.
- Router A, Router B, Router C, and Router D belong to area 10 and are Level-2 routers.
- If load balancing is not enabled, traffic on Router A is transmitted along the optimal route obtained using the SPF calculation. Consequently, traffic on different links is unbalanced. Enabling load balancing on Router A sends traffic to RouterRouter D through RouterRouter B and Router C. This transmission mode relieves the load on the optimal route.

Load balancing can work per-destination or per-packet. For details, see the section "Overview" in *MA5800 Feature Description - IP Routing*.

Administrative Tag

Administrative tags carry administrative information about IP address prefixes. When the cost type is wide, wide-compatible, or compatible and the prefix of the

 Copyright © Huawei Technologies Co., Ltd.

reachable IP address to be advertised by IS-IS has this cost type, IS-IS adds the administrative tag to the reachability type-length-value (TLV) in the prefix. In this manner, the administrative tag is advertised throughout the entire IS-IS area so that routes can be imported or filtered based on the administrative tag.

IS-IS Mesh Group

As defined in IS-IS, a router must flood the received LSP to all neighbors. On a network with multiple connections and point-to-point (P2P) links, this flooding method causes repeated LSP flooding and wastes bandwidth resources.

To address this issue, you can add certain interfaces to a mesh group. These interfaces flood the LSPs received from a group only to interfaces of other groups or interfaces no which no mesh groups are configured. All the interfaces that join a mesh group ensure the synchronization of the LSDBs in the entire network segment using the CSNP and PSNP mechanisms.

11.12.5.2.4 IS-IS Overload

The overload (OL) field of LSPs configured on a device prevents other devices from calculating the routes passing through this device.

If a system fails to store new LSPs for LSDB synchronization, the routes calculated by the system are incorrect. In that case, the system enters the Overload state. The user can configure the device to enter the Overload state when the system lacks sufficient memory. At present, users can set the Overload timer when IS-IS is started and configure whether to delete the leaked routes and whether to advertise the imported routes. A device enters the Overload state after an exception occurs on the device or when it is configured to enter the state.

- If IS-IS enters the Overload state after an exception occurs on the device, the system deletes all imported or leaked routes.
- If IS-IS enters the Overload state based on a user configuration, the system only deletes all imported or leaked routes if configured to do so.

Although LSPs with overload fields are flooded throughout the network, they are ignored in the calculation of the routes passing through the device in the Overload state. Specifically, after the overload field of LSPs is configured on a device, other devices do not count the routes that pass through the device when performing SPF calculation, but the direct routes between the device and other devices are still calculated.

If a device in an IS-IS domain is faulty, routes may be incorrectly calculated across the entire domain. The overload field can be configured for the device to isolate it from the IS-IS network temporarily, which facilitates fault isolation.

11.12.5.2.5 IS-IS Fast Convergence

IS-IS fast convergence is an extended feature of IS-IS implemented to speed up route convergence. It includes the following concepts:

- I-SPF

  Incremental SPF (I-SPF) recalculates only the routes of the changed nodes rather than the routes of all nodes when the network topology changes, which speeds up the calculation of routes.

 Copyright © Huawei Technologies Co., Ltd.

- PRC

  Partial Route Calculation (PRC) calculates only those routes which have changed when the network topology changes.
- LSP fast flooding

  LSP fast flooding speeds up LSP flooding.
- Intelligent timer

  The first timeout period of the timer is fixed. If an event that triggers the timer occurs before the set timer expires, the next timeout period of the timer increases.

  The intelligent timer is applicable to LSP generation and SPF calculation.

I-SPF

In ISO 10589, the Dijkstra algorithm was adopted to calculate routes. When a node changes on the network, the algorithm recalculates all routes. The calculation requires a long time to complete and consumes a significant amount of CPU resources, reducing convergence speed.

I-SPF improves the algorithm. Except for the first time the algorithm is run, only the nodes that have changed rather than all nodes in the network are used in the calculation. The SPT generated using I-SPF is the same as that generated using the previous algorithm. This significantly decreases CPU usage and speeds up network convergence.

PRC

Similar to I-SPF, PRC calculates only routes that have changed. PRC, however, does not calculate the shortest path. It updates routes based on the SPT calculated by I-SPF.

In route calculation, a leaf represents a route, and a node represents a device. If the SPT changes after I-SPF calculation, PRC calculates all the leaves only on the changed node. If the SPT remains unchanged, PRC calculates only the changed leaves.

For example, if IS-IS is enabled on an interface of a node, the SPT calculated by I-SPF remains unchanged. In this case, PRC updates only the routes of this interface, which consumes less CPU resources.

PRC working with I-SPF further improves network convergence performance and replaces the original SPF algorithm.

NOTE

On the MA5800, only I-SPF and PRC are used to calculate IS-IS routes.

LSP Fast Flooding

When an IS-IS device receives new LSPs from other devices, it updates the LSPs in the LSDB and periodically floods the updated LSPs based on a timer. Therefore, the synchronization of all LSDBs is slow.

With lSP fast flooding, when the Router receives LSPs that can trigger route calculation or route update, it floods these LSPs before route calculation occurs,

 Copyright © Huawei Technologies Co., Ltd.

which speeds up network convergence and LSDB synchronization throughout the entire network.

Intelligent Timer

Although the route calculation algorithm is improved, the long interval for triggering route calculation also affects the convergence speed. A millisecond-level timer can shorten the interval. Frequent network changes, however, also consume too much CPU resources. The SPF intelligent timer addresses these problems.

In most cases, an IS-IS network running normally is stable. The frequent changes on a network are rather rare, and IS-IS does not calculate routes frequently. Therefore, a short period (within milliseconds) can be configured as the first interval for route calculation. If the network topology changes frequently, the interval set by the intelligent timer increases with the calculation times to reduce CPU consumption.

The delay for the first SPF calculation is *init-interval*; the delay for the second SPF calculation is *incr-interval*. From the third time on, the delay doubles each time until the delay reaches *max-interval*. After the delay remains *max-interval* three times or the IS-IS process is restarted, the delay decreases to *init-interval*.

The LSP generation intelligent timer is similar to the SPF intelligent timer. When the LSP generation intelligent timer expires, the system generates a new LSP based on the current topology. The original mechanism uses a timer with fixed intervals, which results in slow convergence and high CPU consumption. Therefore, the LSP generation timer is designed as an intelligent timer to respond to emergencies (for example, the interface goes Up or Down) quickly and speed up network convergence. In addition, when the network changes frequently, the interval for the intelligent timer becomes longer to reduce CPU consumption.

11.12.5.2.6 IS-IS LSP Fragment Extension

When the LSPs to be advertised by IS-IS contain a large amount of information, they are advertised in multiple LSP fragments belonging to the same system.

As defined in RFC 3786, virtual system IDs can be configured, and virtual LSPs that carry routing information can be generated for IS-IS.

IS-IS LSP fragment extension allows an IS-IS device to generate more LSP fragments and carry more IS-IS information.

Terms

- Originating system

  The originating system is a device that runs the IS-IS protocol. A single IS-IS process advertises LSPs as virtual devices do, except that the originating system refers to a real IS-IS process.
- Normal system ID

  The normal system ID is the system ID of the originating system.
- Additional system ID

  The additional system ID, assigned by the network administrator, is used to generate additional or extended LSP fragments. A maximum of 256 additional or extended LSP fragments can be generated. Like a normal system ID, an additional system ID must be unique in a routing domain.

 Copyright © Huawei Technologies Co., Ltd.

- Virtual system

  The virtual system, identified by an additional system ID, is used to generate extended LSP fragments. These fragments carry additional system IDs in their LSP IDs.

Principles

IS-IS LSP fragments are identified by the LSP Number field in their LSP IDs. The LSP Number field is 1 byte. Therefore, an IS-ISprocess can generate a maximum of 256 fragments. A 1497-byte LSPcan carry about 30,000 routes. With fragment extension, more information can be carried.

Each system ID represents a virtual system, and each virtual system can generate 256 LSP fragments. In addition, a maximum of another 99 virtual systems can be configured. Therefore, an IS-IS process can generate a maximum of 25,600 LSP fragments.

After a virtual system and fragment extension are configured, an IS-IS device adds the contents that cannot be contained in its LSPs to the LSPs of the virtual system and notifies other devices of the relationship between the virtual system and itself through a special TLV in the LSPs.

IS Alias ID TLV

RFC 3786 defines a special TLV: IS Alias ID TLV.

Table 11-23 IS Alias ID TLV

Field	Length	Description
Type	1 byte	TLV type. If the value is 24, it indicates the IS Alias ID TLV.
Length	1 byte	TLV length.
System ID	6 bytes	System ID.
Pseudonode number	1 byte	Pseudonode number.
sub-TLVs length	1 byte	Length of sub-TLVs.
sub-TLVs	0 to 247 bytes	Sub-TLVs.

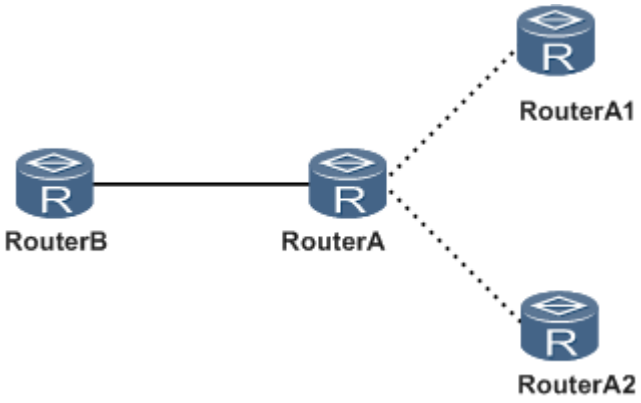
LSPs with fragment number 0 sent by the originating system and virtual system carry IS Alias ID TLVs to indicate the originating system.

Operation Modes

IS-IS devices can use the LSP fragment extension feature in the following modes:

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-95 Networking for IS-IS LSP fragment extension



- Mode-1

  Mode-1 is used when some devices on the network do not support LSP fragment extension.

  In this mode, virtual systems participate in SPF calculation. The originating system advertises LSPs containing information about links to each virtual system and each virtual system advertises LSPs containing information about links to the originating system. In this manner, the virtual systems function the same as the actual devices connected to the originating system on the network.

  Mode-1 is a transitional mode for earlier versions that do not support LSP fragment extension. In the earlier versions, IS-IS cannot identify Alias ID TLVs. Therefore, the LSP sent by a virtual system must look like a common IS-IS LSP.

  The LSP sent by a virtual system contains the same area address and overload bit as those in the common LSP. If the LSPs sent by a virtual system contain TLVs specified in other features, the TLVs must be the same as those in common LSPs.

  LSPs sent by a virtual system carry information of the neighbor (the originating system), and the carried metric is the maximum value minus 1. LSPs sent by the originating system carry information of the neighbor (the virtual system), and the carried metric is 0. This mechanism ensures that the virtual system is a node downstream of the originating system when other devices calculate routes.

  As shown in **Figure 11-95**, Router B does not support LSP fragment extension; Router A supports LSP fragment extension in mode-1; Router A1 and Router A2 are virtual systems of Router A. Router A1 and Router A2 send LSPs carrying partial routing information of Router A. After receiving LSPs from Router A, Router A1, and Router A2, Router B considers there to be three devices at the peer end and calculates routes normally. Because the cost of the route from Router A to Router A1 or Router A2 is 0, the cost of the route from Router B to Router A is equal to that from Router B to Router A1.

- Mode-2

  Mode-2 is used when all the devices on the network support LSP fragment extension. In this mode, virtual systems do not participate in SPF calculation. All the devices on the network know that the LSPs generated by the virtual systems actually belong to the originating system.

 Copyright © Huawei Technologies Co., Ltd.

IS-IS working in mode-2 identifies IS Alias ID TLVs, which are used to calculate the SPT and routes.

As shown in **Figure 11-95**, Router B supports LSP fragment extension, and Router A supports LSP fragment extension in mode-2; Router A1 and Router A2 send LSPs carrying some routing information of Router A. After receiving LSPs from Router A1 and Router A2, Router B obtains IS Alias ID TLV and learns that the originating system of Router A1 and Router A2 is Router A. Router B then considers information advertised by Router A1 and Router A2 to be about Router A.

Whatever the LSP fragment extension mode, LSPs can be resolved. However, if LSP fragment extension is not supported, only LSPs in mode-1 can be resolved.

Table 11-24 Comparison between mode-1 and mode-2

LSP Field	Carried in Mode-1	Carried in Mode-2
IS Alias ID	Yes	Yes
Area	Yes	No
Overload bit	Yes	Yes
IS NBR/IS EXTENDED NBR	Yes	No
Routing	Yes	Yes
ATT bit	Yes, with value 0	Yes, with value 0
P bit	Yes, with value 0	Yes, with value 0

Process

After LSP fragment extension is configured, if information is lost because LSPs overflow, the system restarts the IS-IS process. After being restarted, the originating system loads as much routing information as possible. Any excessive information beyond the forwarding capability of the system is added to the LSPs of the virtual systems for transmission. In addition, if a virtual system with routing information is deleted, the system automatically restarts the IS-IS process.

Usage Scenario

NOTE

If there are non-Huawei devices on the network, LSP fragment extension must be set to mode-1. Otherwise, these devices cannot identify LSPs.

Configuring LSP fragment extension and virtual systems before setting up IS-IS neighbors or importing routes is recommended. If IS-IS neighbors are set up or routes are imported first and the information to be carried exceeds the forwarding capability of 256 fragments before LSP fragment extension and virtual systems are configured, you have to restart the IS-IS process for the configurations to take effect.

 Copyright © Huawei Technologies Co., Ltd.

11.12.5.2.7 IS-IS 3-Way Handshake

IS-IS introduces the 3-way handshake mechanism on P2P links to ensure a reliable data link layer.

Based on ISO 10589, the IS-IS 2-way handshake mechanism uses Hello packets to set up P2P adjacencies between neighboring devices. When a device receives a Hello packet from the other end, it regards the other end as Up and sets up an adjacency with it. However, this mechanism has some serious shortcomings.

When two or more links exist between two devices, an adjacency can still be set up where one link is Down and the other is Up in the same direction. The parameters of the other link are used in SPF calculation. As a result, a device that does not detect any fault along the faulty link will continue trying to forward packets over the link.

The 3-way handshake mechanism resolves these problems on P2P links. In 3-way handshake mode, a device regards a neighbor Up and sets up an adjacency with it only after confirming that the neighbor has received the packet that the device sends.

In addition, the 3-way handshake mechanism uses the 32-bit Extended Local Circuit ID field, which extends the original 8-bit Extended Local Circuit ID field and the limit of only 255 P2P links.

NOTE

By default, the IS-IS 3-way handshake mechanism is implemented on P2P links, as defined in RFC 3373.

11.12.5.2.8 IS-IS NSR

Background

As networks develop, the demand for data, audio, and video services is growing, which imposes increasing requirements on IP network reliability. If an AMB/SMB switchover is performed on a device due to a maintenance operation or a single point of failure, routes may fail to converge, which may result in traffic loss or even a network breakdown. Non-stop routing (NSR) can address this problem and ensure uninterrupted forwarding of key services.

Related Concepts

- High availability (HA): supports data backup between the AMB and SMB.
- Non-stop forwarding (NSF): enables a node to use the GR mechanism to ensure uninterrupted transmission during an AMB/SMB switchover.
- NSR: allows a standby control plane to take over traffic from an active control plane if the active control plane fails, preventing the control planes of neighbors from detecting the fault.
- AMB and SMB: run control plane processes.

Implementation

With IS-IS NSR, IS-IS real-time data is synchronized between the AMB and SMB. After an AMB/SMB switchover is performed on a device, the SMB takes over

 Copyright © Huawei Technologies Co., Ltd.

services from the AMB, and neighbors are unaware of the local fault. After the switchover, the new AMB recovers IS-IS immediately based on the synchronized IS-IS real-time data. Therefore, neighbors are unaware of the switchover as well. IS-IS NSR requires synchronization of the following data:

- All configuration data, such as information about neighbors, timer parameters, and process configurations.
- Dynamic data, such as the interface parameters and state, and information about neighbors and the link state database (LSDB).

Usage Scenario

NSR minimizes the impact of control plane faults and prevents route flapping on networks that require high reliability.

Benefits

NSR improves network reliability and ensures uninterrupted traffic forwarding.

11.12.5.2.9 IS-IS Wide Metric

In the earlier ISO 10589, the largest metric of an interface is 63. TLV type 128 and TLV type 130 contain information about routes, and TLV type 2 contains information about IS-IS neighbors. However, on large-scale networks, the metric range cannot meet the requirements.

As defined in RFC 3784, with IS-IS wide metric, the largest metric of an interface is extended to 16777215, and the largest metric of a route is 4261412864.

After IS-IS wide metric is enabled, TLV type 135 contains information about routes; TLV type 22 contains information about IS-IS neighbors.

- The following lists the TLVs used in narrow mode:
  - IP Internal Reachability TLV: carries routes within an area.
  - IP External Reachability TLV: carries routes outside an area.
  - IS Neighbors TLV: carries information about neighbors.
- The following lists the TLVs used in wide mode:
  - Extended IP Reachability TLV: replaces the earlier IP Reachability TLV and carries information about routes. This TLV expands the range of the route cost to 4 bytes and carries sub-TLVs.
  - IS Extended Neighbors TLV: carries information about neighbors.

NOTE

The metric style can be set to narrow, narrow-compatible, compatible, wide-compatible, or wide mode. **Table 11-25** shows which metric styles are carried in received and sent packets. A device can calculate routes only when it can receive, send, and process corresponding TLVs. Therefore, to ensure correct data forwarding on a network, the proper metric style must be configured for each device on the network.

 Copyright © Huawei Technologies Co., Ltd.

Table 11-25 Metric style carried in received and sent under different metric style configurations

Configured Metric Style	Metric Style Carried in Received Packets	Metric Style Carried in Sent Packets
Narrow	Narrow	Narrow
Narrow-compatible	Narrow and wide	Narrow
Compatible	Narrow and wide	Narrow and wide
Wide-compatible	Narrow and wide	Wide
Wide	Wide	Wide

When the metric style is set to compatible, IS-IS sends the information both in narrow and wide modes.

Process

NOTICE

Once the metric style is changed, the IS-IS process restarts.

- If the metric style carried in sent packets is changed from narrow to wide:

  The information previously carried by TLV type 128, TLV type 130, and TLV type 2 is now carried by TLV type 135 and TLV type 22.
- If the metric style carried in sent packets is changed from wide to narrow:

  The information previously carried by TLV type 135 and TLV type 22 is now carried by TLV type 128, TLV type 130, and TLV type 2.
- If the metric style carried in sent packets is changed from narrow or wide to narrow and wide:

  The information previously carried in narrow or wide mode is now carried by TLV type 128, TLV type 130, TLV type 2, TLV type 135, and TLV type 22.

11.12.5.2.10 IS-IS Authentication

Background

As the Internet develops, more data, voice, and video information are exchanged over the Internet. New services, such as e-commerce, online conferencing and auctions, video on demand, and distance learning, emerge gradually. The new services have high requirements for network security. Carriers need to prevent data packets from being intercepted or modified by attackers or unauthorized users. IS-IS authentication applies to the area or interface where packets need to be protected. Using IS-IS authentication enhances system security and helps carriers provide safe network services.

 Copyright © Huawei Technologies Co., Ltd.

Related Concepts

Authentication Classification

Based on packet types, the authentication is classified as follows:

- Interface authentication: is configured in the interface view to authenticate Level-1 and Level-2 IS-to-IS Hello PDUs (IIHs).
- Area authentication: is configured in the IS-IS process view to authenticate Level-1 CSNPs, PSNPs, and LSPs.
- Routing domain authentication: is configured in the IS-IS process view to authenticate Level-2 CSNPS, PSNPs, and LSPs.

Based on the authentication modes of packets, the authentication is classified into the following types:

- Simple authentication: The authenticated party directly adds the configured password to packets for authentication. This authentication mode provides the lowest password security.
- MD5 authentication: uses the MD5 algorithm to encrypt a password before adding the password to the packet, which improves password security.

NOTICE

MD5 is an insecure encryption algorithm. Therefore, using HMAC-SHA256 or Keychain is recommended.

- Keychain authentication: further improves network security with configurable key chain that changes with time.
- HMAC-SHA256 authentication: uses the HMAC-SHA256 algorithm to encrypt a password before adding the password to the packet, which improves password security.

Implementation

IS-IS authentication encrypts IS-IS packets by adding the authentication field to packets to ensure network security. After receiving IS-IS packets from a remote router, a local router discards the packets if the authentication passwords in the packets are different from the locally configured one. This mechanism protects the local router.

IS-IS provides a type-length-value (TLV) to carry authentication information. The TLV components are as follows:

- Type: indicates the type of a packet, which is 1 byte. The value defined by ISO is 10, while the value defined by IP is 133.
- Length: indicates the length of the authentication TLV, which is 1 byte.
- Value: indicates the authentication information, including authentication type and authenticated password, which ranges from 1 to 254 bytes. The authentication type is 1 byte:
  - 0: reserved

 Copyright © Huawei Technologies Co., Ltd.

- 1: simple authentication
- 3: general authentication, and only HMAC-SHA256 authentication currently
- 54: MD5 authentication
- 255: private authentication

Interface Authentication

Authentication passwords for IIHs are saved on interfaces. The interfaces send authentication packets with the authentication TLV. Interconnected router interfaces must be configured with the same password.

Area Authentication

Every router in an IS-IS area must use the same authentication mode and have the same key chain.

Routing Domain Authentication

Every Level-2 or Level-1-2 router in an IS-IS area must use the same authentication mode and have the same key chain.

For area authentication and routing domain authentication, you can set a router to authenticate SNPs and LSPs separately in the following ways:

- A router sends LSPs and SNPs that carry the authentication TLV and verifies the authentication information of the LSPs and SNPs it receives.
- A router sends LSPs that carry the authentication TLV and verifies the authentication information of the LSPs it receives. The router sends SNPs that carry the authentication TLV and does not verify the authentication information of the SNPs it receives.
- A router sends LSPs that carry the authentication TLV and verifies the authentication information of the LSPs it receives. The router sends SNPs without the authentication TLV and does not verify the authentication information of the SNPs it receives.
- A router sends LSPs and SNPs that carry the authentication TLV but does not verify the authentication information of the LSPs and SNPs it receives.

11.12.5.2.11 IS-IS Control Messages

IS-IS routers implement routing by exchanging control messages. This section describes IS-IS control messages.

IS-IS PDU Formats

Nine types of IS-IS protocol data units (PDUs) are available for processing control information. Each PDU is identified by a 5-digit type code. IS-IS has three major types of PDUs: Hello PDUs, Link State PDUs (LSPs), and Sequence Number PDUs (SNPs). **Table 11-26** shows the mapping between PDUs and type values.

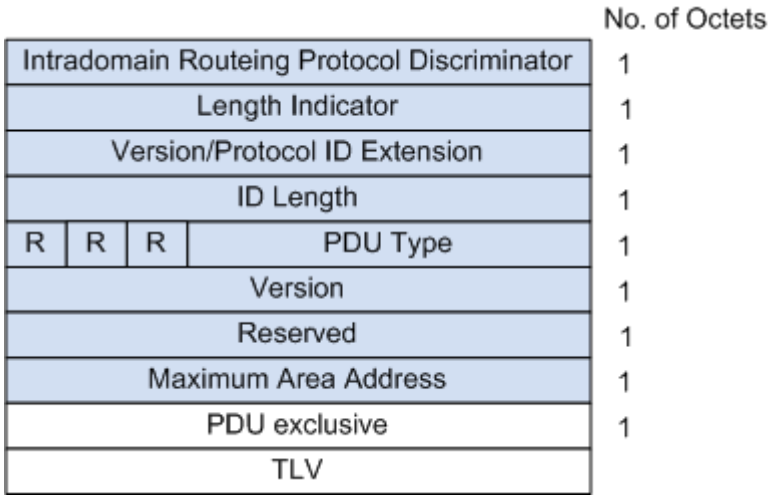
 Copyright © Huawei Technologies Co., Ltd.

Table 11-26 Mapping between PDUs and type values

PDU Type	Acronym	Type Value
Level-1 LAN IS-IS Hello PDU	L1 LAN IIH	15
Level-2 LAN IS-IS Hello PDU	L2 LAN IIH	16
Point-to-Point IS-IS Hello PDU	P2P IIH	17
Level-1 Link State PDU	L1 LSP	18
Level-2 Link State PDU	L2 LSP	20
Level-1 Complete Sequence Numbers PDU	L1 CSNP	24
Level-2 Complete Sequence Numbers PDU	L2 CSNP	25
Level-1 Partial Sequence Numbers PDU	L1 PSNP	26
Level-2 Partial Sequence Numbers PDU	L2 PSNP	27

The first eight bytes in all IS-IS PDUs are public. **Figure 11-96** shows the IS-IS PDU format.

Figure 11-96 IS-IS PDU format



The main fields are as follows:

- Intradomain Routing Protocol Discriminator: network layer protocol identifier assigned to IS-IS, which is 0x83.
- Length Indicator: length of the fixed header, in bytes.
- ID Length: length of the system ID of network service access point (NSAP) addresses or NETs in this routing domain.
- PDU Type: type of a PDU. For details, see **Table 11-26**.

 Copyright © Huawei Technologies Co., Ltd.

- Maximum Area Address: maximum number of area addresses supported by an IS-IS area. The value **0** indicates that a maximum of three area addresses are supported by this IS-IS area.
- Type/Length/Value (TLV): encoding type that features high efficiency and expansibility. Each type of PDU contains a different TLV. **Table 11-27** shows the mapping between TLV codes and PDU types.

Table 11-27 Mapping between TLV codes and PDU types

TLV Code	TLV Code Name	PDU Type
1	Area Addresses	IIH, LSP
2	IS Neighbors (LSP)	LSP
4	Partition Designated Level2 IS	L2 LSP
6	IS Neighbors (MAC Address)	LAN IIH
7	IS Neighbors (SNPA Address)	LAN IIH
8	Padding	IIH
9	LSP Entries	SNP
10	Authentication Information	IIH, LSP, or SNP
128	IP Internal Reachability Information	LSP
129	Protocols Supported	IIH or LSP
130	IP External Reachability Information	L2 LSP
131	Inter-Domain Routing Protocol Information	L2 LSP
132	IP Interface Address	IIH or LSP

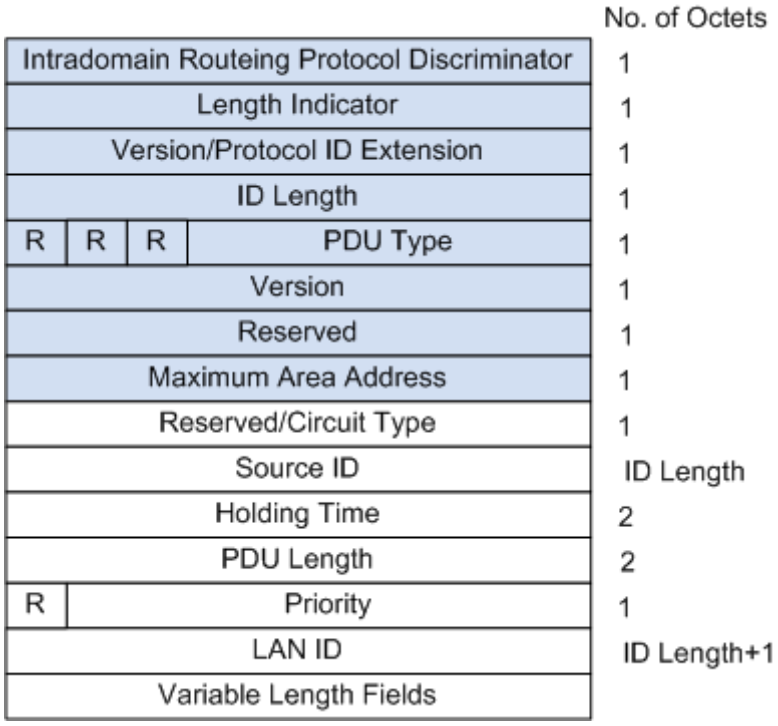
Hello Packet Format

Hello packets, also called the IS-to-IS Hello PDUs (IIHs), are used to set up and maintain neighbor relationships. Level-1 LAN IIHs are applied to the Level-1 routers on broadcast LANs. Level-2 LAN IIHs are applied to the Level-2 routers on broadcast LANs. P2P IIHs are applied to non-broadcast networks. IIHs in different networks have different formats.

- LAN IIHs: **Figure 11-97** shows the format of IIHs on a broadcast network.

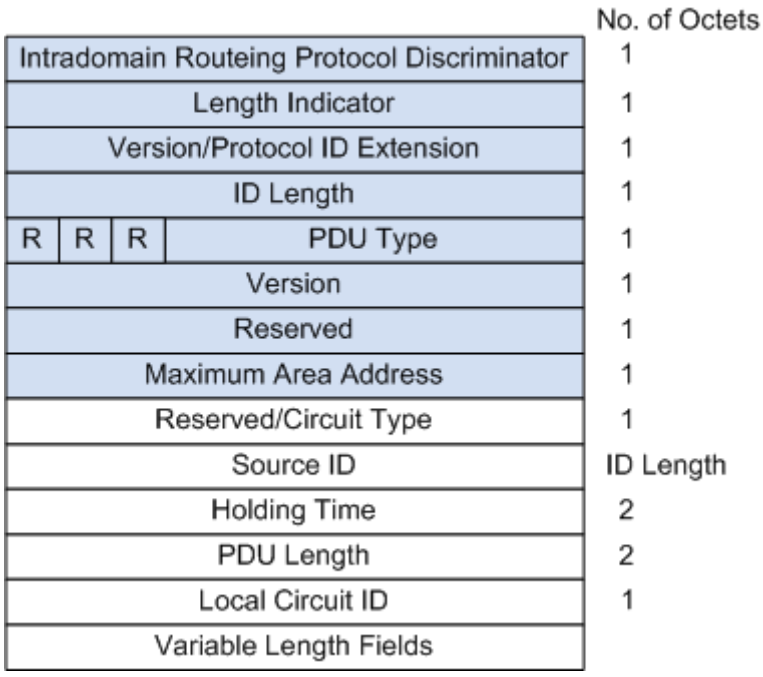
 Copyright © Huawei Technologies Co., Ltd.

Figure 11-97 Level-1/Level-2 LAN IIH format



- P2P IIHs: **Figure 11-98** shows the format of IIHs on a P2P network.

Figure 11-98 P2P IIH format



As shown in **Figure 11-98**, most fields in a P2P IIH are the same as those in a LAN IIH. The P2P IIH does not have the priority and LAN ID fields but has a local circuit ID field. The local circuit ID indicates the local link ID.

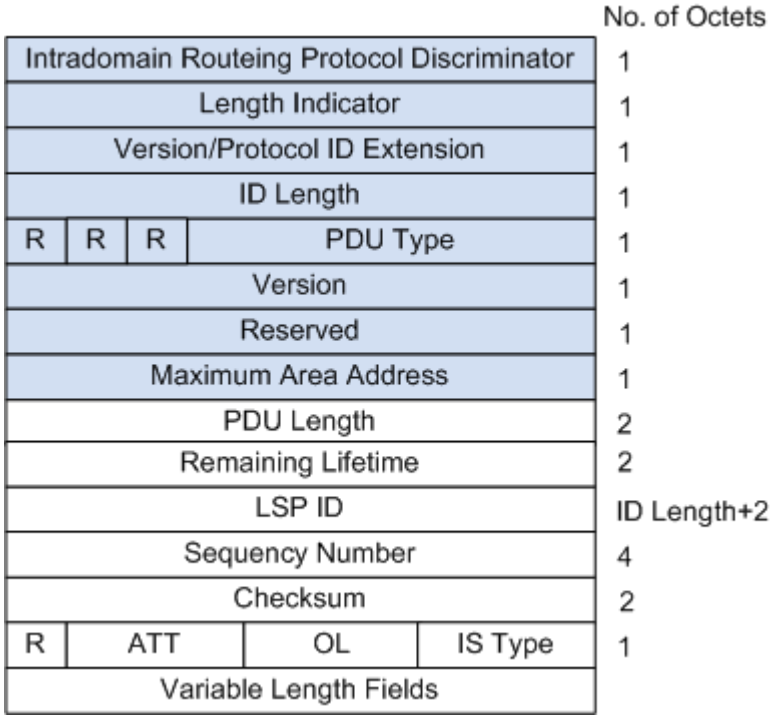
 Copyright © Huawei Technologies Co., Ltd.

LSP Format

LSPs are used to exchange link-state information. There are two types of LSPs: Level-1 and Level-2. Level-1 IS-IS transmits Level-1 LSPs. Level-2 IS-IS transmits Level-2 LSPs. Level-1-2 IS-IS can transmit both Level-1 and Level-2 LSPs.

Level-1 and Level-2 LSPs have the same format, as shown in **Figure 11-99**.

Figure 11-99 Level-1 or Level-2 LSP



The main fields are as follows:

- ATT: Attached bit

  ATT is generated by a Level-1-2 router to identify whether the originating router is connected to other areas. When a Level-1 router receives a Level-1 LSP with ATT as 1 from a Level-1-2 router, the Level-1 router generates a default route destined for the Level-1-2 router so that data can be transmitted to other areas.

  Although ATT is defined in both the Level-1 LSP and Level-2 LSP, it is set only in the Level-1 LSP only by the Level-1-2 router.

- OL: LSDB overload

  LSPs with the overload bit are still flooded on networks, but the LSPs are not used when routes that pass through a device configured with the overload bit are calculated. That is, after a device is configured with the overload bit, other devices ignore the device when performing the SPF calculation except for the direct routes of the device.

- IS Type: type of the IS-IS generating the LSP

  IS Type is used to specify whether the IS-IS type is Level-1 or Level-2 IS-IS. The value **01** indicates Level-1; the value **11** indicates Level-2.

 Copyright © Huawei Technologies Co., Ltd.

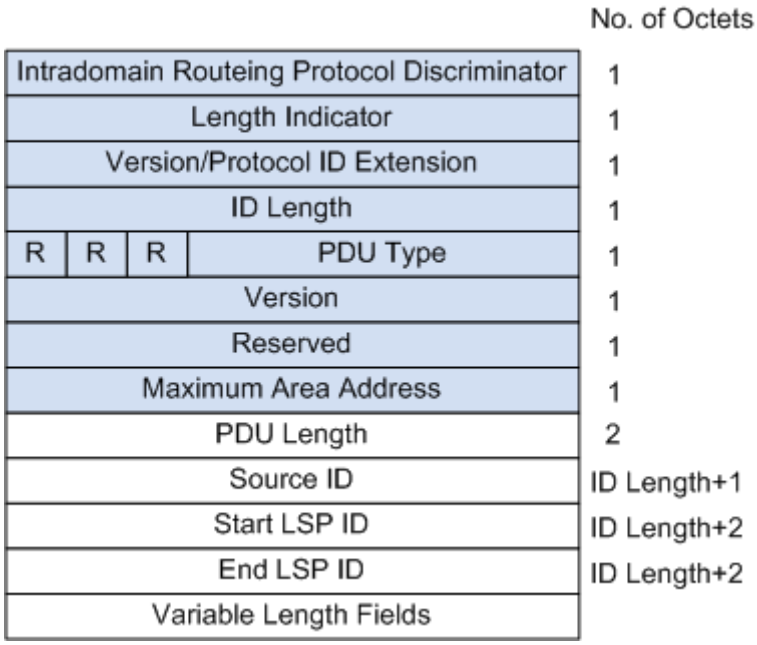
SNP Format

SNPs describe the LSPs in all or some of the databases and are used to synchronize and maintain all LSDBs. SNPs consist of complete SNPs (CSNPs) and partial SNPs (PSNPs).

- CSNPs carry summaries of all LSPs in LSDBs, which ensures LSDB synchronization between neighboring routers. On a broadcast network, the designated intermediate system (DIS) sends CSNPs at an interval. The default interval is 10 seconds. On a P2P link, neighboring devices send CSNPs only when a neighbor relationship is established for the first time.

  Figure 11-100 shows the CSNP format.

  Figure 11-100 Level-1/Level-2 CSNP format

  The main fields are as follows:

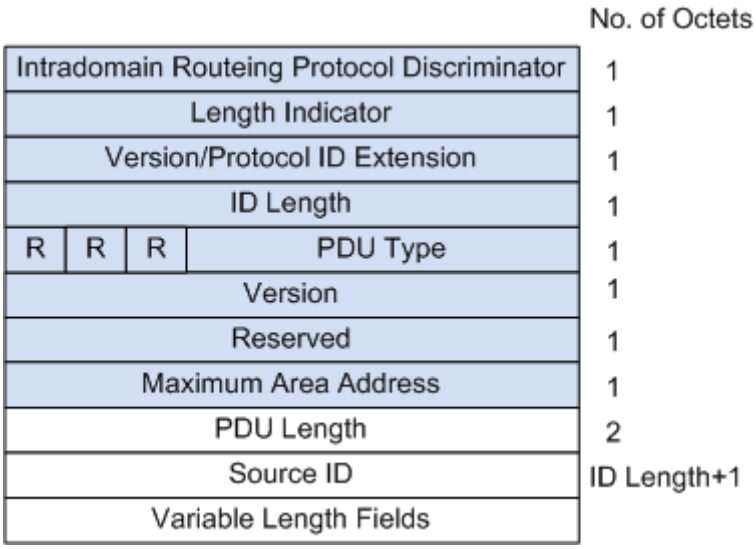
  - Source ID: system ID of the router that sends SNPs
  - Start LSP ID: ID of the first LSP in a CSNP
  - End LSP ID: ID of the last LSP in a CSNP

- PSNPs list only the sequence numbers of recently received LSPs. A PSNP can acknowledge multiple LSPs at a time. If an LSDB is not updated, PSNPs are also used to request a new LSP from a neighbor.

  Figure 11-101 shows the PSNP format.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-101 Level-1/Level-2 PSNP format



11.12.5.2.12 IS-IS for IPv6

RFC 5308 (Routing IPv6 with IS-IS) released by the IETF defines two new TLVs that can support IPv6 routes and a new Network Layer Protocol Identifier (NLPID), which ensures that IS-IS can process and calculate IPv6 routes.

The two new TLVs are as follows:

- IPv6 Reachability

  The IPv6 Reachability TLV indicates the reachability of a network by specifying the route prefix and metric. The type value is 236 (0xEC).

- IPv6 Interface Address

  The IPv6 Interface Address TLV is similar to the IP interface address TLV of IPv4 in function, except that it changes the original 32-bit IPv4 address to a 128-bit IPv6 address. The type value is 232 (0xE8).

The NLPID is an 8-bit field that identifies network layer protocol packets. The NLPID of IPv6 is 142 (0x8E). If an IS-IS router supports IPv6, it advertises routing information through the NLPID value.

11.12.5.2.13 IS-IS Auto FRR

Background

Some services, such as voice over IP (VoIP) and online video services, are delay-sensitive. Nevertheless, if a primary link fails, IS-IS must perform multiple operations, including detecting the fault, updating the link-state advertisement (LSA), flooding the LSA, calculating routes, and delivering forward information base (FIB) entries before switching traffic to a new link. This process takes a much longer time than the delay-sensitive services can tolerate. To solve this problem, IS—IS Auto FRR is developed.

IS—IS auto fast reroute (FRR) is a kind of dynamic IP FRR. To implement this feature, Interior Gateway Protocol (IGP) calculates the backup paths beforehand by using the link status database of the entire network and then stores the backup

 Copyright © Huawei Technologies Co., Ltd.

paths in the forwarding table to minimize service interruption duration in case of a link fault.

LFA

A device uses each neighbor that can be used as a backup next hop as a root node, uses the shortest path first (SPF) algorithm to calculate the shortest path from each root node to the destination, and then uses the inequalities defined in RFC 5286 to calculate the backup link that is loop-free and has the lowest cost.

The link costs meet the inequality: Distance_opt(N, D) < Distance_opt(N, S) + Distance_opt(S, D). Distance_opt(X, Y) indicates the shortest link from X to Y. S stands for a source node, E for the faulty node, N for a node along a backup link, and D for a destination node.

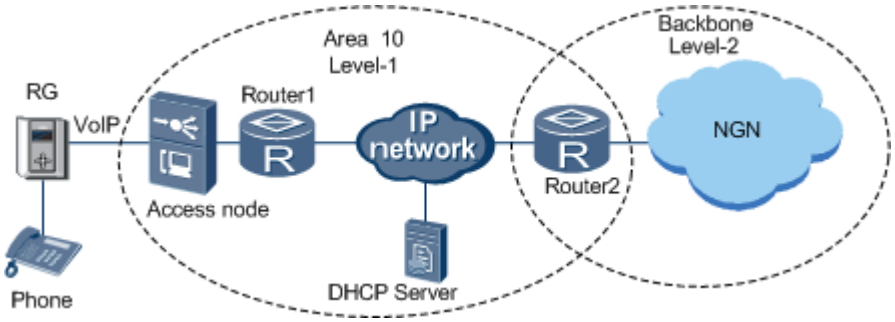
11.12.5.3 Configuration Example of IS-IS

This operation enables the corresponding device configured data to run the IS-IS protocol on the access device.

Service Requirements

- The access device forwards the access VoIP service through the Layer 3 interface to the NGN network.
- The access device obtains the routes of the NGN networking through the IS-IS protocol. The area ID of the Level-2 router differs from the area ID of the Level-1-2 router to which the Level-2 router connects.

Figure 11-102 Example network for configuring IS-IS



Data Plan

Table 11-28 provides the data plan for configuring IS-IS.

Table 11-28 Data plan for configuring IS-IS

Item	Data
Access node	IS-IS process ID: 1

 Copyright © Huawei Technologies Co., Ltd.

Item	Data
	NET (Network entity title): 10.0000.0000.0001.00, where: ● Area ID: 10 ● System ID: 0000.0000.0001 ● Level: Level-1 ● Host name: MA5800
	IS-IS interface: ● Port number: 0/9/0 ● VLAN ID: 20 ● IP address: 10.5.5.5/16
Router1	IS-IS process ID: 1
	NET (Network entity title): 10.0000.0000.0002.00, where: ● Area ID: 10 ● System ID: 0000.0000.0002 ● Level: Level-1 ● Host name: Router1
	IS-IS interface: 1/0/0 IP address: 10.8.8.8/16
Router2	IS-IS process ID: 1
	NET (Network entity title): 10.0000.0000.0005.00, where: ● Area ID: 10 ● System ID: 0000.0000.0005 ● Level: Level-1-2 ● Host name: Router2
	IS-IS interface: 1/0/0 IP address: 10.9.9.9/16

Procedure

- Configure IS-IS on the access node.
  a. Configure the Layer 3 interface.

```
huawei(config)#vlan 20 standard
huawei(config)#port vlan 20 0/9 0
huawei(config)#interface vlanif 20
huawei(config-if-vlanif20)#ip address 10.5.5.5 16
huawei(config-if-vlanif20)#quit
```

  b. Start the IS-IS process.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#isis 1
huawei(config-isis-1)#
```

c. Configure the NET.

```
huawei(config-isis-1)#network-entity 10.0000.0000.0001.00
```

d. Configure the router level.

```
huawei(config-isis-1)#is-level level-1
```

e. Configure the local host name.

```
huawei(config-isis-1)#is-name MA5800
huawei(config-isis-1)#quit
```

f. Enable the IS-IS function on an interface.

```
huawei(config)#interface vlanif 20
huawei(config-if-vlanif20)#isis enable 1
```

- Configure IS-IS on Router1.

  The process of configuring IS-IS on Router1 is similar to that of configuring IS-IS on the access node. The details are not provided in this chapter.

- Configure IS-IS on Router2.

  The process of configuring IS-IS on Router2 is similar to that of configuring IS-IS on the access node. The details are not provided in this chapter.

----End

Result

- Run the **display isis lsdb** command and you can query the IS-IS LSDB.
- Run the **display isis route** command and you can query the IS-IS route. The routing table of the Level-1 router should have a default route, and the next hop should be the Level-1-2 router. The Level-2 router should have the routes to all the Level-1 routers and the Level-2 routers.

Configuration File

```
vlan 20 standard
port vlan 20 0/9 0
interface vlanif 20
ip address 10.5.5.5 16
quit
isis 1
network-entity 10.0000.0000.0001.00
is-level level-1
is-name MA5800
quit
interface vlanif 20
isis enable 1
```

11.12.5.4 Configuration Example of IS-IS for IPv6

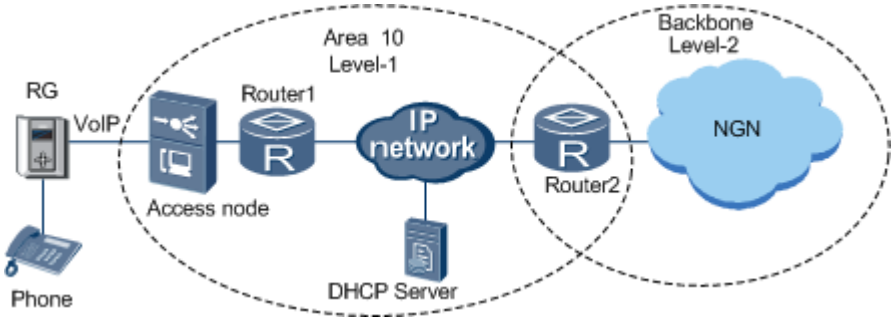
This operation enables the corresponding device configured data to run the IS-IS protocol on the access device in the IPv6 network.

Service Requirements

- The access device forwards the access VoIP service through the Layer 3 interface to the NGN network.
- The access device obtains the routes of the NGN networking through the IS-IS protocol.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-103 Example network for configuring IS-IS for IPv6



Data Plan

Table 11-29 provides the data plan for configuring IS-IS.

Table 11-29 Data plan for configuring IS-IS

Item	Data
Access node	IS-IS process ID: 1
	NET (Network entity title): 10.0000.0000.0001.00, where: • Area ID: 10 • System ID: 0000.0000.0001 • Level: Level-1 • Host name: MA5800
	IS-IS interface: • Port number: 0/9/0 • VLAN ID: 20 • IP address: 2001:db8:1::1/64

Procedure

- Configure IS-IS on the access node.
  a. Configure the Layer 3 interface.

```
huawei(config)#vlan 20 standard
huawei(config)#port vlan 20 0/9 0
huawei(config)#interface vlanif 20
huawei(config-if-vlanif20)#ipv6 enable
huawei(config-if-vlanif20)#ipv6 address 2001:db8:1::1 64
huawei(config-if-vlanif20)#quit
```

  b. Start the IS-IS process, and enable IPv6 for the process.

```
huawei(config)#isis 1
huawei(config-isis-1)#ipv6 enable
```

  c. Configure the NET.

```
huawei(config-isis-1)#network-entity 10.0000.0000.0001.00
```

 Copyright © Huawei Technologies Co., Ltd.

d. Configure the router level.

```
huawei(config-isis-1)#is-level level-1
```

e. Configure the local host name.

```
huawei(config-isis-1)#is-name MA5800
huawei(config-isis-1)#quit
```

f. Enable the IS-IS for IPv6 function on an interface.

```
huawei(config)#interface vlanif 20
huawei(config-if-vlanif20)#isis ipv6 enable 1
huawei(config-if-vlanif20)#quit
```

- Configure IS-IS for IPv6 on Router1.

  The process of configuring IS-IS on Router1 is similar to that of configuring IS-IS on the access node. The details are not provided in this chapter.

- Configure IS-IS for IPv6 on Router2.

  The process of configuring IS-IS on Router2 is similar to that of configuring IS-IS on the access node. The details are not provided in this chapter.

----End

Result

- Run the **display isis lsdb** command and you can query the IS-IS LSDB.
- Run the **display isis route** command and you can query the IS-IS route. The routing table of the Level-1 router should have a default route, and the next hop should be the Level-1-2 router. The Level-2 router should have the routes to all the Level-1 routers and the Level-2 routers.

Configuration File

```
vlan 20 standard
port vlan 20 0/9 0
interface vlanif 20
ipv6 enable
ipv6 address 2001:db8:1::1 64
quit
isis 1
ipv6 enable
network-entity 10.0000.0000.0001.00
is-level level-1
is-name MA5800
quit
interface vlanif 20
isis ipv6 enable 1
quit
```

11.12.5.5 Example for Configuring IS-IS Auto FRR

IS-IS auto fast reroute (FRR) is applicable to services that are sensitive to packet loss and latency. Configuring IS-IS auto FRR implements fast link switching, which greatly improves the reliability of an IS-IS network.

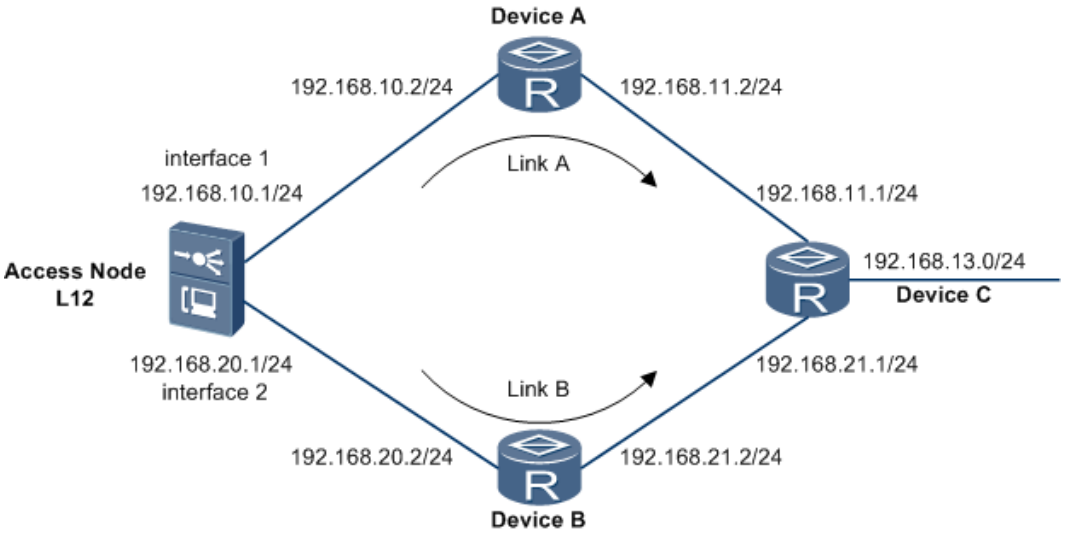
Service Requirements

When a fault occurs on the network, IS-IS auto FRR fast switches traffic to the backup link without waiting for route convergence, reducing the switching time.

As shown in **Figure 11-104**,

- The 4 routers run IS-IS to interconnect IP networks.
- The 4 routers are level-1-2 routers.
- When link A fails, the MA5800 fast switches traffic to backup link B.

Figure 11-104 Networking of configuring IS-IS auto FRR



Data Plan

Table 11-30 describes data plan for configuring IP IS-IS auto FRR.

Table 11-30 Data plan for configuring IS-IS auto FRR

Item	Data
MA5800 port 1	IP address of the port: 192.168.10.1 24
	VLAN ID: 2
	Cost: 10
MA5800 port 2	IP address of the port: 192.168.20.1 24
	VLAN ID: 3
	Cost: 30

Prerequisite

Before you configure IS-IS Auto FRR, complete the following tasks:

- Configuring link layer protocol parameters and IP addresses for ports to ensure that the link layer protocol on the ports is Up
- Configuring basic IP IS-IS functions

 Copyright © Huawei Technologies Co., Ltd.

Procedure

Step 1 Perform configuration on the MA5800.

1. Configure the IP address of the Layer 3 interface on the MA5800.

```
huawei(config)#vlan 2 smart
huawei(config)#port vlan 2 0/9 0
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 192.168.10.1 24
huawei(config-if-vlanif2)#quit
huawei(config)#vlan 3 smart
huawei(config)#port vlan 3 0/9 1
huawei(config)#interface vlanif 3
huawei(config-if-vlanif3)#ip address 192.168.20.1 24
huawei(config-if-vlanif3)#quit
```

2. Configure basic IS-IS functions.

```
huawei(config)#isis 1
huawei(config-isis-1)#is-level level-1-2
huawei(config-isis-1)#network-entity 10.0000.0000.0001.00
huawei(config-isis-1)#quit
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#isis enable 1
huawei(config-if-vlanif2)#quit
huawei(config)#interface vlanif 3
huawei(config-if-vlanif3)#isis enable 1
huawei(config-if-vlanif3)#quit
```

3. Set the port cost of the MA5800 to 30.

```
huawei(config)#interface vlan 3
huawei(config-if-vlanif3)#isis cost 30
huawei(config-if-vlanif3)#quit
```

4. Enable IS-IS auto FRR on the MA5800.

```
huawei(config)#isis
huawei(config-isis-1)#frr
huawei(config-isis-1-ipv6-frr)#loop-free-alternate
huawei(config-isis-1-ipv6-frr)#quit
huawei(config-isis-1)#quit
```

5. Save the data.

```
huawei(config)#save
```

----End

Result

Run the **display isis route** command to query the information about primary/secondary routes configured.

Configuration File

```
vlan 2 smart
port vlan 2 0/9 0
interface vlanif 2
ip address 192.168.10.1 24
quit
vlan 3 smart
port vlan 3 0/9 1
interface vlanif 3
ip address 192.168.20.1 24
quit
isis 1
is-level level-1-2
network-entity 10.0000.0000.0001.00
quit
```

 Copyright © Huawei Technologies Co., Ltd.

```
interface vlanif 2
isis enable 1
quit
interface vlanif 3
isis enable 1
quit
interface vlan 3
isis cost 30
quit
isis
frr
loop-free-alternate
quit
quit
save
```

11.12.5.6 Example for Configuring IPv6 IS-IS Auto FRR

IPv6 IS-IS auto fast reroute (FRR) is applicable to services that are sensitive to packet loss and latency. Configuring IPv6 IS-IS auto FRR implements fast link switching, which greatly improves the reliability of an IS-IS network.

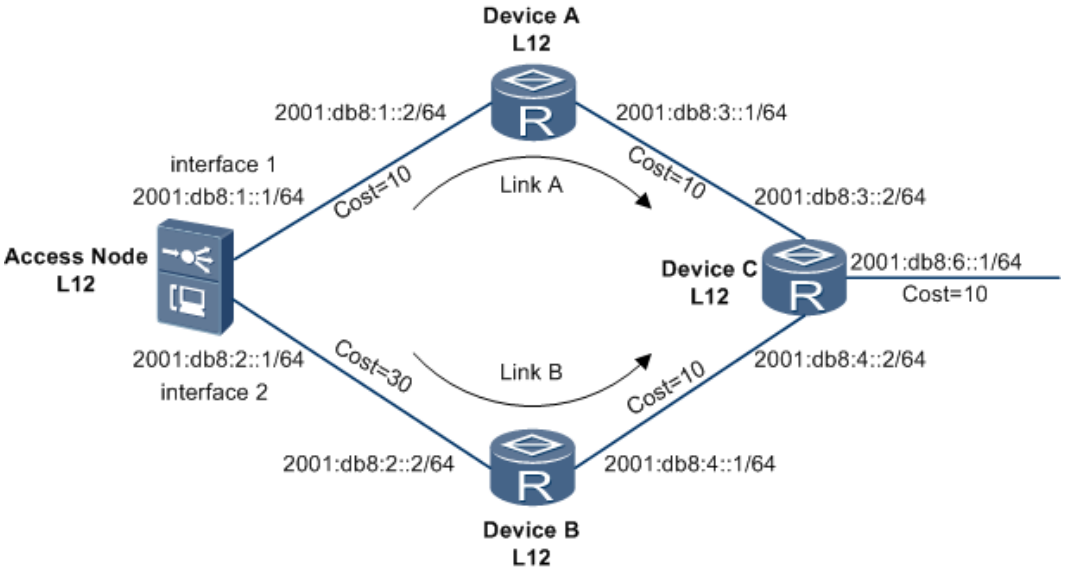
Service Requirements

When a fault occurs on the network, IPv6 IS-IS auto FRR fast switches traffic to the backup link without waiting for route convergence, reducing the switching time.

As shown in **Figure 11-105**,

- The 4 routers run IS-IS to interconnect IPv6 networks.
- The 4 routers are level-1-2 routers.
- When link A fails, the MA5800 fast switches traffic to backup link B.

Figure 11-105 Networking of configuring IPv6 IS-IS auto FRR



Data Plan

Table 11-31 describes data plan for configuring IPv6 IS-IS auto FRR.

Table 11-31 Data plan for configuring IPv6 IS-IS auto FRR

Item	Data
MA5800 port 1	IPv6 address of the port: 2001:db8:1::1 64
	VLAN ID: 2
	Cost: 10
MA5800 port 2	IPv6 address of the port: 2001:db8:2::1 64
	VLAN ID: 3
	Cost: 30

Prerequisite

Before you configure IPv6 IS-IS Auto FRR, complete the following tasks:

- Configuring link layer protocol parameters and IPv6 addresses for ports to ensure that the link layer protocol on the ports is Up
- Configuring basic IPv6 IS-IS functions

Procedure

Step 1 Perform configuration on the MA5800.

1. Configure the IPv6 address of the Layer 3 interface on the MA5800.

```
huawei(config)#vlan 2 smart
huawei(config)#port vlan 2 0/9 0
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ipv6 enable
huawei(config-if-vlanif2)#ipv6 address 2001:db8:1::1 64
huawei(config-if-vlanif2)#quit
huawei(config)#vlan 3 smart
huawei(config)#port vlan 3 0/9 1
huawei(config)#interface vlanif 3
huawei(config-if-vlanif3)#ipv6 enable
huawei(config-if-vlanif3)#ipv6 address 2001:db8:2::1 64
huawei(config-if-vlanif3)#quit
```

2. Configure basic IS-IS IPv6 functions.

```
huawei(config)#isis 1
huawei(config-isis-1)#is-level level-1-2
huawei(config-isis-1)#network-entity 10.0000.0000.0001.00
huawei(config-isis-1)#quit
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#isis ipv6 enable 1
huawei(config-if-vlanif2)#quit
huawei(config)#interface vlanif 3
huawei(config-if-vlanif3)#isis ipv6 enable 1
huawei(config-if-vlanif3)#quit
```

3. Set the port cost of the MA5800 to 30.

```
huawei(config)#interface vlan 3
huawei(config-if-vlanif3)#isis ipv6 cost 30
huawei(config-if-vlanif3)#quit
```

 Copyright © Huawei Technologies Co., Ltd.

4. Enable IPv6 IS-IS auto FRR on the MA5800.

```
huawei(config)#isis
huawei(config-isis-1)#ipv6 frr
huawei(config-isis-1-ipv6-frr)#loop-free-alternate
huawei(config-isis-1-ipv6-frr)#quit
huawei(config-isis-1)#quit
```

5. Save the data.

```
huawei(config)#save
```

----End

Result

Run the **display isis route** command to query the information about primary/secondary routes configured.

Configuration File

```
vlan 2 smart
port vlan 2 0/9 0
interface vlanif 2
ipv6 enable
ipv6 address 2001:db8:1::1 64
quit
vlan 3 smart
port vlan 3 0/9 1
interface vlanif 3
ipv6 enable
ipv6 address 2001:db8:2::1 64
quit
isis 1
is-level level-1-2
network-entity 10.0000.0000.0001.00
quit
interface vlanif 2
isis ipv6 enable 1
quit
interface vlanif 3
isis ipv6 enable 1
quit
interface vlan 3
isis ipv6 cost 30
quit
isis
ipv6 frr
loop-free-alternate
quit
quit
save
```

11.12.5.7 IS-IS Reference Standards and Protocols

Table 11-32 The following table lists the references.

Document NO.	Document Name	Remarks
ISO 10589	ISO IS-IS Routing Protocol	-
RFC 1142	OSI IS-IS Intra-domain Routing Protocol	-

 Copyright © Huawei Technologies Co., Ltd.

Document NO.	Document Name	Remarks
RFC 1195	Use of OSI IS-IS for Routing in TCP/IP and Dual Environments	Multiple authentication passwords cannot be configured.
RFC 2763	Dynamic Hostname Exchange Mechanism for IS-IS	-
RFC 2966	Domain-wide Prefix Distribution with Two-Level IS-IS	-
RFC 2973	IS-IS Mesh Groups	-
RFC 3277	IS-IS Transient Blackhole Avoidance	-
RFC 3359	Summary of all Reserved Type, Length and Value (TLV) Codepoints in IS-IS	-
RFC 3373	Three-Way Handshake for IS-IS Point-to-Point Adjacencies	-
RFC 3567	IS-IS Cryptographic Authentication	-
RFC 3719	Recommendations for Interoperable Networks using IS-IS	-
RFC 3786	Extension of the Number of IS-IS LSP Fragments Beyond the 256 Limit	-
RFC 3787	Recommendations for Interoperable IP Networks using IS-IS	-
RFC 3906	Calculation of Interior Gateway Protocol (IGP) Routes	-
RFC 4444	Management Information Base for IS-IS	Only some features are supported.
RFC 5130	A Policy Control Mechanism in IS-IS Using Administrative Tags	-
RFC 5286	Basic Specification for IP Fast Reroute: Loop-Free Alternates	-
RFC 5301	Dynamic Hostname Exchange Mechanism for IS-IS	Same as RFC 2763.
RFC 5302	Domain-Wide Prefix Distribution with Two-Level IS-IS	Same as RFC 2966.
RFC 5303	Three-Way Handshake for IS-IS Point-to-Point Adjacencies	Same as RFC 3373.
RFC 5304	IS-IS Cryptographic Authentication	Same as RFC 3567.

 Copyright © Huawei Technologies Co., Ltd.

Document NO.	Document Name	Remarks
RFC 5309	Point-to-Point Operation over LAN in Link State Routing Protocols	-
draft-ietf-isis-igp-p2p-over-lan-06	Point-to-point operations over LAN in link-state routing protocols	-
draft-ietf-isis-3way-03	Three-Way Handshake for IS-IS Point-to-Point Adjacencies	-
draft-ietf-rtgwg-ipfrr-framework-11	IP Fast Reroute Framework	-

11.12.6 OSPF

11.12.6.1 Introduction to OSPF

Definition

Open Shortest Path First (OSPF) is a link-state Interior Gateway Protocol (IGP) developed by the Internet Engineering Task Force (IETF).

OSPF version 2 (OSPFv2), which is defined in RFC 2328, is intended for IPv4. OSPF version 3 (OSPFv3), which is defined in RFC 2740, is intended for IPv6.

NOTE

In this document, OSPF refers to OSPFv2, unless otherwise stated.

Purpose

Before the emergence of OSPF, the Routing Information Protocol (RIP) was widely used as an IGP on networks. RIP is a distance-vector routing protocol, and therefore has slow convergence, routing loops, and poor scalability. OSPF was developed to replace RIP. Typical IGPs include RIP, OSPF, and Intermediate System to Intermediate System (IS-IS). **Table 11-33** describes differences among the three typical IGPs.

Table 11-33 Differences among the three typical IGPs

Item	RIP	OSPF	IS-IS
Protocol type	IP layer protocol	IP layer protocol	Link layer protocol

 Copyright © Huawei Technologies Co., Ltd.

Item	RIP	OSPF	IS-IS
Application scope	Applies to small networks with simple architectures, such as campus networks.	Applies to medium-sized networks with several hundred Routers supported, such as enterprise networks.	Applies to large networks, such as Internet service provider (ISP) networks.
Routing algorithm	Uses a distance-vector algorithm and exchanges routing information over the User Datagram Protocol (UDP).	Uses the shortest path first (SPF) algorithm to generate a shortest path tree (SPT) based on the network topology, calculates shortest paths to all destinations, and exchanges routing information over IP.	Uses the SPF algorithm to generate an SPT based on the network topology, calculates shortest paths to all destinations, and exchanges routing information over IP. The SPF algorithm runs separately in Level-1 and Level-2 databases.
Route convergence speed	Slow	Less than 1 second	Less than 1 second
Scalability	Not supported	Supported by partitioning a network into areas	Supported by defining Router levels

Benefits

OSPF offers the following benefits:

- Wide application scope: OSPF applies to medium-sized networks with several hundred Routers supported, such as enterprise networks.
- Network masks supported: OSPF packets can carry masks, and therefore the packet length is not limited by natural IP masks. OSPF can process variable length subnet masks (VLSMs).
- Fast convergence: When the network topology changes, OSPF immediately sends link state update (LSU) packets to synchronize the changes to the link state databases (LSDBs) of all routers in an autonomous system (AS).
- Loop-free: OSPF uses the SPF algorithm to calculate loop-free routes based on the collected link status.
- Area partition: OSPF allows an AS to be partitioned into areas to simplify management. Routing information transmitted between areas is further summarized, which reduces network bandwidth consumption.

 Copyright © Huawei Technologies Co., Ltd.

- Equal-cost route: OSPF allows multiple equal-cost routes to the same destination.
- Routing hierarchy: OSPF uses four route types: intra-area routes, inter-area routes, Type 1 external routes, and Type 2 external routes, which are listed in descending order of priority.
- Authentication: OSPF supports area- and interface-based packet authentication, which ensures packet exchange security.
- Multicast: OSPF uses multicast addresses to send packets on certain types of links, which minimizes the impact on other devices.

11.12.6.2 Principles

11.12.6.2.1 Basic Concepts of OSPF

This section describes the basic Open Shortest Path First (OSPF) concepts.

Router ID

A router ID is a 32-bit unsigned integer, which identifies a Router in an autonomous system (AS). A router ID must exist before a Router runs OSPF.

A router ID can be obtained in two modes:

- In manual mode, a router ID needs to be manually configured.
- Automatically set by the router.

  If no router ID has been configured, the router automatically obtains a router ID using the following methods in descending order of priority. The selection sequence is as follows:

  a. The router preferentially selects the largest IP address from its loopback interface addresses as the router ID.

  b. If no loopback interface has been configured, the router selects the largest IP address from its interface IP addresses as the router ID.

A router can obtain a router ID again only after a router ID is reconfigured for the router or an OSPF router ID is reconfigured and the OSPF process restarts.

Area

When a large number of Routers run OSPF, link state databases (LSDBs) become very large and require a large amount of storage space. Large LSDBs also complicate shortest path first (SPF) computation and are computationally intensive for the Routers. Network expansion causes the network topology to change, which results in route flapping and frequent OSPF packet transmission. When a large number of OSPF packets are transmitted on the network, bandwidth usage efficiency decreases. Each change in the network topology causes all Routers on the network to recalculate routes.

OSPF resolves this problem by partitioning an AS into different areas. An area is regarded as a logical group, and each group is identified by an area ID. A Router, not a link, resides at the border of an area. A network segment or link can belong only to one area. An area must be specified for each OSPF interface.

 Copyright © Huawei Technologies Co., Ltd.

OSPF areas include common areas, stub areas, and not-so-stubby areas (NSSAs), as described in **Table 11-34**.

Table 11-34 OSPF areas

Area Type	Function	Notes
Common area	By default, OSPF areas are defined as common areas. Common areas include: • Standard area: transmits intra-area, inter-area, and external routes. • Backbone area: connects to all other OSPF areas and transmits inter-area routes. The backbone area is represented by area 0. Routes between non-backbone areas must be forwarded through the backbone area.	• The backbone area must have all its devices connected. • All non-backbone areas must remain connected to the backbone area.
Stub area	A stub area is a non-backbone area with only one ABR and generally resides at the border of an AS. The area border router (ABR) in a stub area does not transmit received AS external routes, which significantly decreases the number of entries in the routing table on the ABR and the amount of routing information to be transmitted. To ensure the reachability of AS external routes, the ABR in the stub area generates a default route and advertises the route to non-ABRs in the stub area. A totally stub area allows only intra-area routes and ABR-advertised Type 3 link state advertisements (LSAs) carrying a default route to be advertised within the area.	• The backbone area cannot be configured as a stub area. • An autonomous system boundary router (ASBR) cannot exist in a stub area. Therefore, AS external routes cannot be advertised within the stub area. • A virtual link cannot pass through a stub area.
NSSA	An NSSA is similar to a stub area. An NSSA does not advertise **Type5 LSA**s, but can import AS external routes. ASBRs in an NSSA generate **Type7 LSA**s to carry the information about the AS external routes. The Type 7 LSAs are advertised only within the NSSA. When the Type 7 LSAs reach an ABR in the NSSA, the ABR translates the Type 7 LSAs into Type 5 LSAs and floods them to the entire AS. A totally NSSA area allows only intra-area routes to be advertised within the area.	• ABRs in an NSSA advertise Type 7 LSAs carrying a default route within the NSSA. All inter-area routes are advertised by ABRs. • A virtual link cannot pass through an NSSA.

 Copyright © Huawei Technologies Co., Ltd.

Router Type

Routers are classified as internal routers, ABRs, backbone routers, or ASBRs by location in an AS. **Figure 11-106** and **Table 11-35** shows the four router types.

Figure 11-106 Router type layout

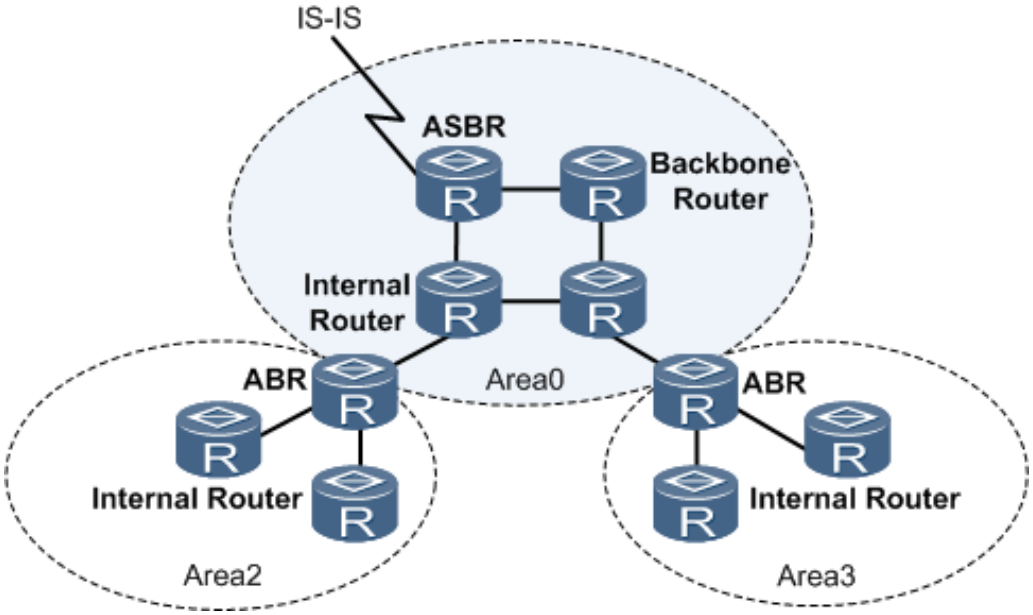


Table 11-35 Router type description

Router Type	Description
Internal router	All interfaces of an internal router belong to the same OSPF area.
ABR	An ABR can belong to two or more areas, one of which must be a **backbone area**. An ABR connects the backbone area and non-backbone areas. An ABR can be physically or logically connected to the backbone area.
Backbone router	At least one interface on a backbone router belongs to the backbone area. Internal routers in the backbone area and all ABRs are backbone routers.
ASBR	An ASBR exchanges routing information with other ASs. An ASBR may not reside at the boundary of an AS. An ASBR can be an internal router or an ABR.

 Copyright © Huawei Technologies Co., Ltd.

LSA

OSPF encapsulates routing information into LSAs for transmission. **Table 11-36** describes LSAs and their functions.

Table 11-36 LSAs and their functions

LSA Type	Function
Router-LSA (Type 1)	Describes the link status and cost of a Router. Router-LSAs are generated by a Router and advertised within the area to which the Router belongs.
Network-LSA (Type 2)	Describes the link status of all routers on the local network segment. Network-LSAs are generated by a designated router (DR) and advertised within the area to which the DR belongs.
Network-summary-LSA (Type 3)	Describes routes on a network segment. Network-summary-LSAs are generated by an **ABR** and are advertised within the non-totally stub area or NSSA.
ASBR-summary-LSA (Type 4)	Describes routes to an ASBR in an area. ASBR-summary-LSAs are generated by an **ABR** and are advertised to the areas except the area to which the ASBR belongs.
AS-external-LSA (Type 5)	Describes AS external routes, which are advertised to all areas except **stub areas** and **NSSAs**. AS-external-LSAs are generated by an **ASBR**.
NSSA-LSA (Type 7)	Describes AS external routes. NSSA-LSAs are generated by an **ASBR** and advertised only within NSSAs.
Opaque-LSA (Type 9/ Type 10/Type 11)	Provides a general mechanism for OSPF extension. Different types of LSAs are described as follows: • Type 9 LSAs are advertised only on the network segment where the interface advertising the LSAs resides. Grace LSAs used in graceful restart (GR) are Type 9 LSAs. • Type 10 LSAs are advertised within an OSPF area. LSAs that are used to support traffic engineering (TE) are Type 10 LSAs. • Type 11 LSAs are advertised within an AS but have not been used in practice.

Packet Type

OSPF packets are encapsulated into IP packets. The OSPF protocol number is 89. OSPF packets are classified as Hello, Database Description (DD), Link State

 Copyright © Huawei Technologies Co., Ltd.

Request (LSR), Link State Update (LSU), or Link State Acknowledgment (LSAck) packets, as described in **Table 11-37**.

Table 11-37 OSPF packets and their functions

Packet Type	Function
Hello packet	Hello packets are periodically sent to discover and maintain OSPF neighbor relationships.
DD packet	A DD packet contains the summaries of LSAs in the local LSDB. DD packets are used for LSDB synchronization between two Routers.
LSR packet	LSR packets are sent to OSPF neighbors to request required LSAs. A Router sends LSR packets to its OSPF neighbor only after DD packets have been successfully exchanged.
LSU packet	LSU packets are used to transmit required LSAs to OSPF neighbors.
LSAck packet	LSAck packets are used to acknowledge received LSAs.

Route Type

Intra-area and inter-area routes describe the network structure of an AS. AS external routes describe how to select routes to destinations outside an AS. AS external routes imported by OSPF are classified as Type 1 or Type 2 external routes.

Table 11-38 describes OSPF routes in descending order of priority.

Table 11-38 OSPF routes

Route Type	Description
Intra-area route	-
Inter-area route	-
Type 1 external route	Because a Type 1 external route has high reliability, its cost equals the cost of an AS internal route. Cost of a Type 1 external route = Cost of the route from a Router to an ASBR + Cost of the route from the ASBR to the destination

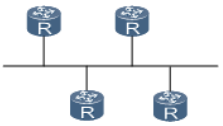
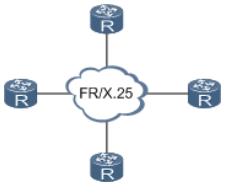
 Copyright © Huawei Technologies Co., Ltd.

Route Type	Description
Type 2 external route	Because a Type 2 external route has low reliability, its cost is considered to be much greater than the cost of any internal route to an ASBR. Cost of a Type 2 external route = Cost of the route from an ASBR to the destination

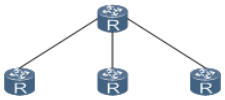
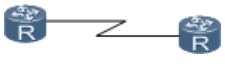
Network Type

Networks are classified as broadcast, non-broadcast multiple access (NBMA), point-to-multipoint (P2MP), or point-to-point (P2P) networks by link layer protocol, as described in **Table 11-39**.

Table 11-39 OSPF network classification

Network Type	Link Layer Protocol	Packet Transmission Mode	Graph
Broadcast	• Ethernet • FDDI	• **Hello** and **LSAck** packets are transmitted by an OSPF device using the address 224.0.0.5 in multicast mode. **LSU** packets are first transmitted in multicast mode and retransmitted in unicast mode. When an OSPF DR is used, the OSPF DR transmits packets using the address 224.0.0.6 in multicast mode. • **DD** and **LSR** packets are transmitted in unicast mode.	
NBMA	X.25	Hello, DD, LSR, LSU, and LSAck packets are transmitted in unicast mode.	

 Copyright © Huawei Technologies Co., Ltd.

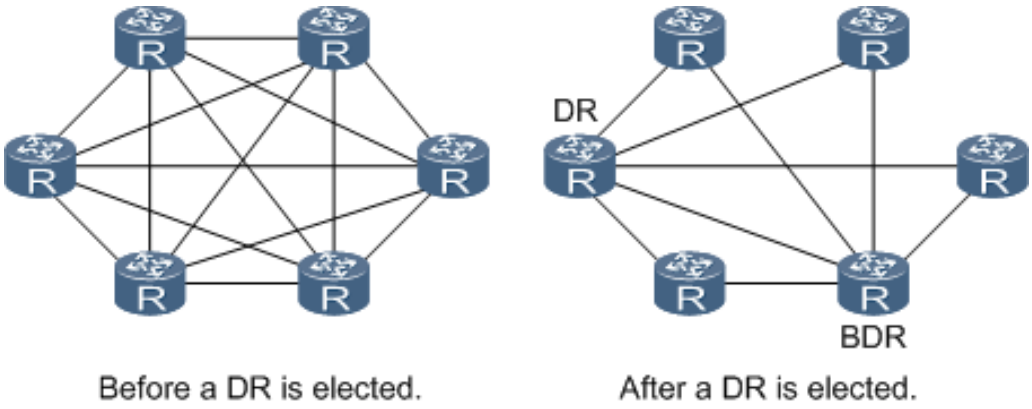
Network Type	Link Layer Protocol	Packet Transmission Mode	Graph
P2MP	Regardless of the link layer protocol, OSPF does not default the network type to P2MP. The network type must be manually changed to P2MP. The common practice is to change a non-fully connected NBMA network to a P2MP network.	• Hello packets are transmitted using the address 224.0.0.5 in multicast mode. • DD, LSR, LSU, and LSAck packets are transmitted in unicast mode.	
P2P	• PPP • HDLC • LAPB	• Hello, DD, LSR, LSU, and LSAck packets are transmitted using the address 224.0.0.5 in multicast mode. • LSU packets are retransmitted in multicast mode.	

 Copyright © Huawei Technologies Co., Ltd.

DR and BDR

On broadcast or NBMA networks, any two Routers need to exchange routing information. As shown in **Figure 11-107**, n routers are deployed on the network. n x (n - 1)/2 adjacencies must be established. Any route change on a Router is transmitted to other Routers, which wastes bandwidth resources. OSPF resolves this problem by defining a DR and a backup designated router (BDR). After a DR is elected, all Routers send routing information only to the DR. Then the DR broadcasts LSAs. Routers other than the DR and BDR are called DR others. The DR others establish only adjacencies with the DR and BDR and not with each other. This process reduces the number of adjacencies established between Routers on broadcast or NBMA networks.

Figure 11-107 Network topologies before and after a DR election



If the original DR fails, Routers must reelect a DR and the Routers except the new DR must synchronize routing information to the new DR. This process is lengthy, which may cause incorrect route calculations. A BDR is used to shorten the process. The BDR is a backup for a DR. A BDR is elected together with a DR. The BDR establishes adjacencies with all Routers on the network segment and exchanges routing information with them. When the DR fails, the BDR immediately becomes a new DR. The Routers need to reelect a new BDR, but this process does not affect route calculations.

The DR priority of a Router interface determines its qualification for DR and BDR elections. The Router interfaces with their DR priorities greater than 0 are eligible. Each Router adds the elected DR to a Hello packet and sends it to other Routers on the network segment. When both Router interfaces on the same network segment declare that they are DRs, the Router interface with a higher DR priority is elected as a DR. If the two Router interfaces have the same DR priority, the Router interface with a larger router ID is elected as a DR.

OSPF Multi-process

OSPF supports multi-process. Multiple OSPF processes can independently run on the same Router. Route exchange between different OSPF processes is similar to route exchange between different routing protocols. A Router's interface can belong only to one OSPF process.

 Copyright © Huawei Technologies Co., Ltd.

A typical application of OSPF multi-process is that OSPF runs between provider edges (PEs) and customer edges (CEs) on virtual private networks (VPNs) and OSPF is also used as an Interior Gateway Protocol (IGP) on the VPN backbone network. The two OSPF processes on the PEs are independent of each other.

OSPF Default Route

A default route is the route whose destination address and mask are both all 0s. When no exactly matching route is discovered, a Router uses a default route to forward packets.

A default route generally applies to the following scenarios:

- An ABR in an area advertises Type 3 LSAs carrying a default route within the area. The Routers in the area use the received default route to forward inter-area packets.
- An ASBR in an AS advertises Type 5 or Type 7 LSAs carrying a default route within the AS. The Routers in the AS use the received default route to forward AS external packets.

OSPF routes are hierarchically managed. The priority of the default route carried in Type 3 LSAs is higher than the priority of the default route carried in Type 5 or Type 7 LSAs.

The basic principles for advertising LSAs carrying a default route are as follows:

- A Router in an area can advertise LSAs carrying a default route only when the Router has an interface connected to a device outside the area.
- If a Router has advertised LSAs carrying a default route, the Router no longer learns the same type of LSA advertised by other Routers, which carry a default route. That is, the Router uses only the LSAs advertised by itself to calculate routes. The LSAs advertised by other Routers are still saved in the LSDB.
- If a Router must use a route to advertise LSAs carrying an external default route, the route cannot be a route learned by the local OSPF process. A Router in an area uses an external default route to forward packets outside the area. If the next hops of routes in the area are Routers in the area, packets cannot be forwarded outside the area.
- Before a Router advertises a default route, it checks whether a neighbor in the full state is present in area 0. The Router advertises a default route only when a neighbor in the full state is present in area 0. If no such a neighbor exists, the backbone area cannot forward packets and advertising a default route is meaningless.

Table 11-40 describes the principles for advertising default routes in different areas.

 Copyright © Huawei Technologies Co., Ltd.

Table 11-40 Principles for advertising default routes in different areas

Area Type	Advertisement Principles
Common area	By default, a Router in a common area does not generate a default route. After being configured to do so, an ASBR generates a Type 5 LSA carrying a default route. The Router then advertises the default route in the entire AS. If no default route is generated on the ASBR, the Router does not advertise a default route.
Stub area	Type 5 LSAs cannot be advertised within a stub area. A Router in the stub area must learn AS external routes from an ABR. The ABR automatically generates a Type 3 LSA carrying a default route and advertises it within the entire stub area. Then the Router can learn AS external routes from the ABR.
Totally stub area	Neither Type 3 nor Type 5 LSAs can be advertised within a totally stub area. A Router in the totally stub area must learn AS external and inter-area routes from an ABR. After you configure a totally stub area, an ABR automatically generates a Type 3 LSA carrying a default route and advertises it within the entire totally stub area. Then the Router can learn AS external and inter-area routes from the ABR.
NSSA	A small number of AS external routes learned from the ASBR in an NSSA can be imported to the NSSA. Type 5 LSAs cannot be advertised within the NSSA. The ABR automatically generates a Type 7 LSA carrying a default route and advertises it within the entire NSSA. A small number of AS external routes can be learned from the ASBR in the NSSA, and other inter-area routes can be learned from the ABR in the NSSA. Manual configurations must be performed on the ASBR to enable the ASBR to generate a Type 7 LSA carrying a default route and advertise the LSA within the entire NSSA. An ABR does not translate Type 7 LSAs carrying a default route into Type 5 LSAs carrying a default route or flood them to the entire AS.
Totally NSSA	Neither Type 3 nor Type 5 LSAs can be advertised within a totally NSSA. A Router in the totally NSSA must learn AS external routes from an ABR. The ABR automatically generates Type 3 and Type7 LSAs carrying a default route and advertises them to the entire totally NSSA. Then AS external and inter-area routes can be advertised within the totally NSSA.

11.12.6.2.2 Basic Principles of OSPF

OSPF route calculation involves the following processes:

 Copyright © Huawei Technologies Co., Ltd.

1. The **adjacency establishment** process is as follows:
   a. The local device sends Hello packets through an interface to a remote device to establish a **neighbor relationship**.
   b. The local and remote devices negotiate a master/slave relationship and exchange Database Description (DD) packets.
   c. The local and remote devices exchange link state advertisements (LSAs) to synchronize their link state databases (LSDBs).

   The **adjacency** is established.
2. **Route calculation**

   OSPF uses the shortest path first (SPF) algorithm to calculate routes, resulting in fast route convergence.

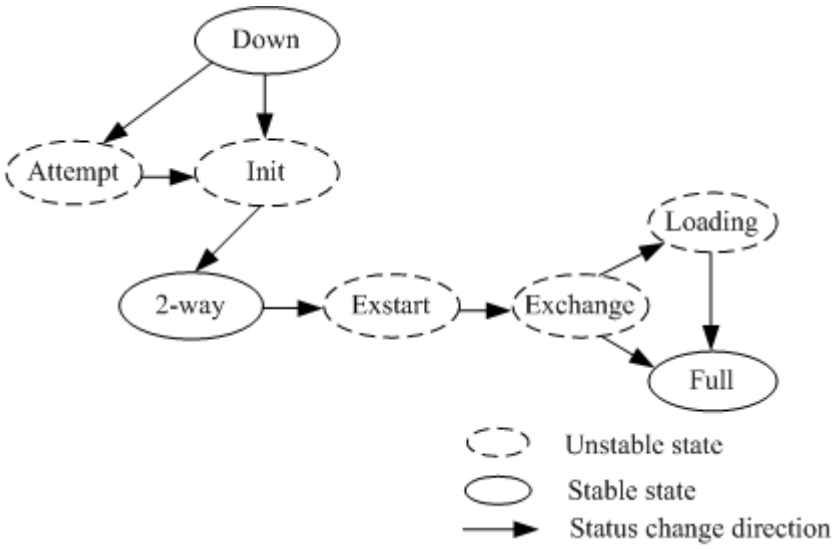
OSPF Neighbor States

To exchange routing information on an OSPF network, neighbor Routers must establish adjacencies. The differences between neighbor relationships and adjacencies are described as follows:

- Neighbor relationship: After the local Router starts, it uses an OSPF interface to send a Hello packet to the remote Router. After the remote Router receives the packet, it checks whether the parameters carried in the packet are consistent with its own parameters. If the parameters carried in the packet are consistent with its own parameters, the local and remote Routers establish a neighbor relationship.
- Adjacency: After the local and remote Routers establish a neighbor relationship, they exchange DD packets and LSAs to establish an adjacency.

OSPF has eight neighbor states: Down, Attempt, Init, 2-way, Exstart, Exchange, Loading, and Full, as shown in **Figure 11-108**. Down, 2-way, and Full are stable states. Attempt, Init, Exstart, Exchange, and Loading are unstable states, which last only several minutes.

Figure 11-108 OSPF neighbor states



 Copyright © Huawei Technologies Co., Ltd.

Table 11-41 OSPF neighbor states and their meanings

OSPF Neighbor State	Meaning
Down	This is the initial state of a neighbor conversation. This state indicates that a Router has not received any Hello packets from its neighbors within a dead interval.
Attempt	In the Attempt state, a Router periodically sends Hello packets to manually configured neighbors. **NOTE** This state applies only to non-broadcast multiple access (NBMA) interfaces.
Init	This state indicates that a Router has received Hello packets from its neighbors but the neighbors did not receive Hello packets from the Router.
2-way	This state indicates that a Router has received Hello packets from its neighbors and **neighbor relationships** have been established between the Routers. If no adjacency needs to be established, the neighbors remain in the 2-way state. If adjacencies need to be established, the neighbors enter the Exstart state.
Exstart	In the Exstart state, Routers establish a master/slave relationship to ensure that DD packets are sequentially exchanged.
Exchange	In the Exchange state, Routers exchange DD packets. A Router uses a DD packet to describe its own LSDB and sends the packet to its neighbors.
Loading	In the Loading state, a Router sends Link State Request (LSR) packets to its neighbors to request their LSAs for LSDB synchronization.
Full	In the Full state, a Router establishes **adjacencies** with its neighbors and all LSDBs have been synchronized.

NOTE

The neighbor state of the local Router may be different from that of the remote Router. For example, the neighbor state of the local Router is Full, but the neighbor state of the remote Router is Loading.

Adjacency Establishment

Adjacencies can be established in either of the following situations:

- Two Routers have established a neighbor relationship and communicate for the first time.
- The designated router (DR) or backup designated router (BDR) on a network segment changes.

 Copyright © Huawei Technologies Co., Ltd.

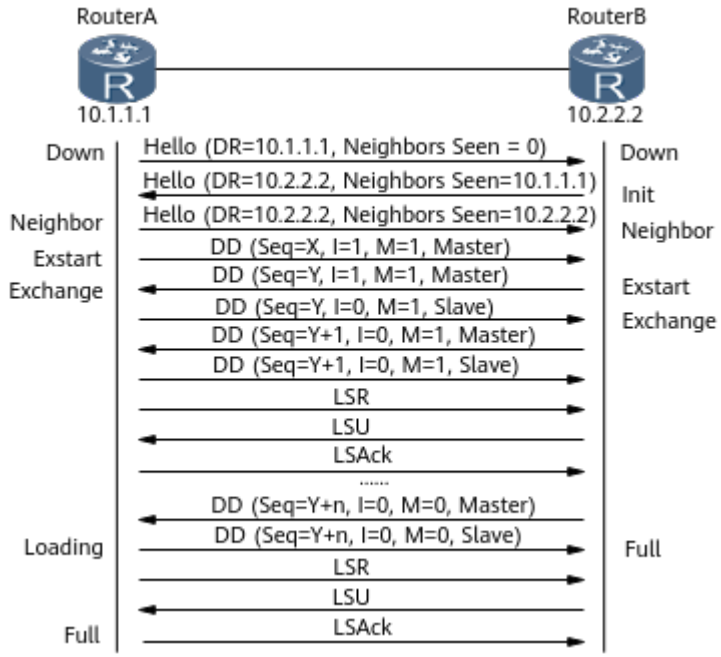
The adjacency establishment process is different on different networks.

Adjacency establishment on a broadcast network

On a broadcast network, the DR and BDR establish adjacencies with each Router on the same network segment, but DR others establish only neighbor relationships.

Figure 11-109 shows the adjacency establishment process on a broadcast network.

Figure 11-109 Adjacency establishment process on a broadcast network



The adjacency establishment process on a broadcast network is as follows:

1. Neighbor relationship establishment
   a. Router A uses the multicast address 224.0.0.5 to send a Hello packet to Router B through the OSPF interface connected to a broadcast network. The packet carries the DR field of 10.1.1.1 (ID of Router A) and the Neighbors Seen field of 0. A neighbor Router has not been discovered, and Router A regards itself as a DR.
   b. After Router B receives the packet, it returns a Hello packet to Router A. The returned packet carries the DR field of 10.2.2.2 (ID of Router B) and the Neighbors Seen field of 10.1.1.1 (Router A's router ID). Router A has been discovered but its router ID is less than that of Router B, and therefore Router B regards itself as a DR. Then Router B's status changes to Init.
   c. After Router A receives the packet, Router A's status changes to Exstart.

   NOTE

   The following procedures are not performed for DR others on a broadcast network.
2. Master/Slave negotiation and DD packet exchange

 Copyright © Huawei Technologies Co., Ltd.

a. Router A sends a DD packet and claims it as the master (MS = 1), and sets the sequence number (Seq) to x. I field: The value 1 indicates that the packet is the first DD packet, which is used to negotiate a master/slave relationship and does not carry LSA summaries. M field: The value 1 indicates that the packet is not the last DD packet.

   To improve transmission efficiency, Router A and Router B determine which LSAs in each other's LSDB need to be updated. If one party determines that an LSA of the other party is already in its own LSDB, it does not send an LSR packet for updating the LSA to the other party. To achieve the preceding purpose, Router A and Router B first send DD packets, which carry summaries of LSAs in their own LSDBs. Each summary identifies an LSA. To ensure packet transmission reliability, a master/slave relationship must be determined during DD packet exchange. One party serving as a master uses the Seq field to define a sequence number. The master increases the sequence number by one each time it sends a DD packet. When the other party serving as a slave sends a DD packet, it adds the sequence number carried in the last DD packet received from the master to the Seq field of the packet.

b. After Router B receives the DD packet, Router B's status changes to Exstart and Router B returns a DD packet to Router A. The returned packet does not carry LSA summaries. Because Router B's router ID is greater than Router A's router ID, Router B declares itself a master and sets the Seq field to *y*.

c. After Router A receives the DD packet, it agrees that Router B is a master and Router A's status changes to Exchange. Then Router A sends a DD packet to Router B to transmit LSA summaries. The packet carries the Seq field of *y* and the MS field of 0. The value 0 indicates that Router A declares itself a slave.

d. After Router B receives the packet, Router B's status changes to Exchange and Router B sends a new DD packet containing its own LSA summaries to Router A. The value of the Seq field carried in the new DD packet is changed to *y* + 1.

Router A uses the same sequence number as Router B to confirm that it has received DD packets from Router B. Router B uses the sequence number plus one to confirm that it has received DD packets from Router A. When Router B sends the last DD packet, it sets the M field of the packet to 0.

3. LSDB synchronization

   a. After Router A receives the last DD packet, it finds that many LSAs in Router B's LSDB do not exist in its own LSDB, so Router A's status changes to Loading. After Router B receives the last DD packet from Router A, Router B's status directly changes to Full, because Router B's LSDB already contains all LSAs of Router A.

   b. Router A sends an LSR packet for updating LSAs to Router B. Router B returns an LSU packet to Router A. After Router A receives the packet, it sends an LSAck packet for acknowledgement.

   The preceding procedures continue until the LSAs in Router A's LSDB are the same as those in Router B's LSDB. Router A's status changes to Full. After Router A and Router B exchange DD packets and update all LSAs, they establish an adjacency.

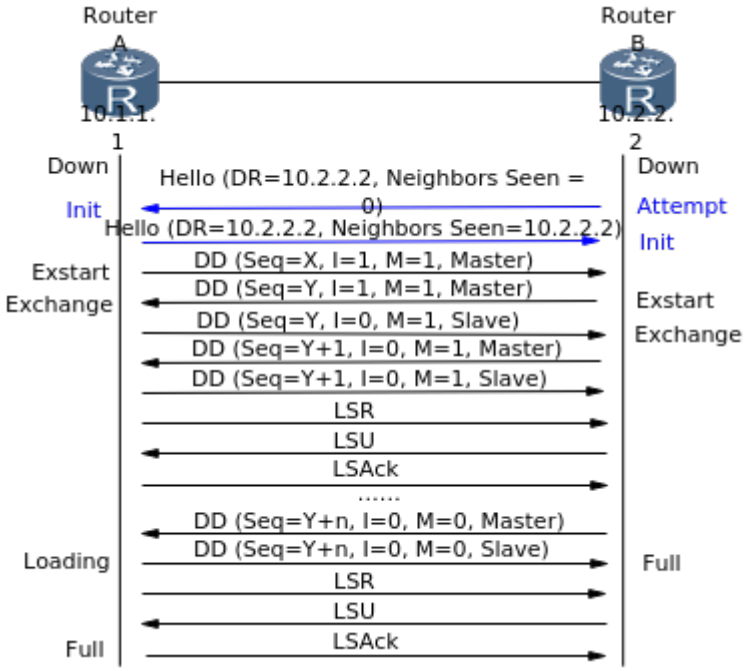
Adjacency establishment on an NBMA network

 Copyright © Huawei Technologies Co., Ltd.

The adjacency establishment process on an NBMA network is similar to that on a broadcast network. The blue part shown in **Figure 11-110** highlights the differences from a broadcast network.

On an NBMA network, all Routers establish adjacencies only with the DR and BDR.

Figure 11-110 Adjacency establishment process on an NBMA network



The adjacency establishment process on an NBMA network is as follows:

1. Neighbor relationship establishment
   a. After Router B sends a Hello packet to a Down interface of Router A, Router B's status changes to Attempt. The packet carries the DR field of 2.2.2.2 (ID of Router B) and the Neighbors Seen field of 0. A neighbor Router has not been discovered, and Router B regards itself as a DR.
   b. After Router A receives the packet, Router A's status changes to Init and Router A returns a Hello packet. The returned packet carries the DR and Neighbors Seen fields of 2.2.2.2. Router B has been discovered but its router ID is greater than that of Router A, and therefore Router A agrees that Router B is a DR.

   NOTE

   The following procedures are not performed for DR others on an NBMA network.
2. The procedures for negotiating a master/slave relationship and exchanging DD packets on an NBMA network are the same as those on a broadcast network.
3. The procedure for synchronizing LSDBs (LSA request, LSA transmission, and LSA response) on an NBMA network is the same as that on a broadcast network.

 Copyright © Huawei Technologies Co., Ltd.

Adjacency establishment on a point-to-point (P2P)/point-to-multipoint (P2MP) network

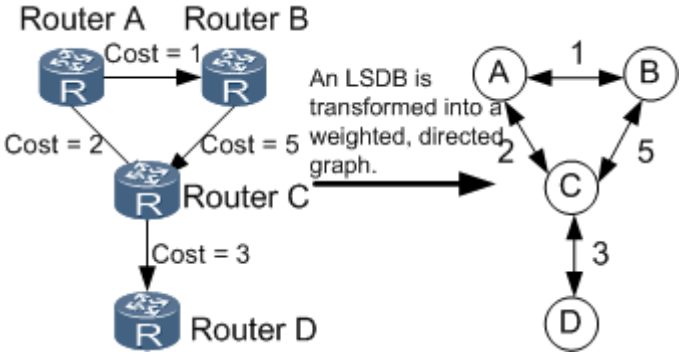
The adjacency establishment process on a P2P/P2MP network is similar to that on a broadcast network. On a P2P/P2MP network, however, no DR or BDR needs to be elected and DD packets are transmitted in multicast mode.

Route Calculation

OSPF uses the shortest path first (SPF) algorithm to calculate routes, resulting in fast route convergence.

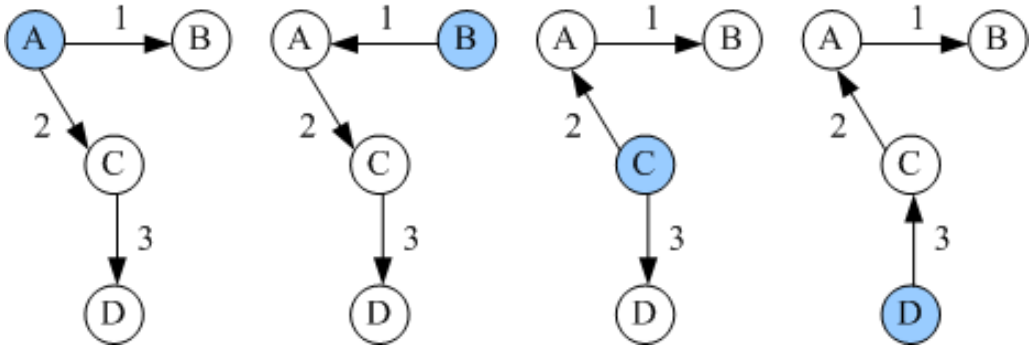
OSPF uses an LSA to describe the network topology. A Type 1 LSA describes the attributes of a link between Routers. A Router transforms its LSDB into a weighted, directed graph, which reflects the topology of the entire AS. All Routers in the same area have the same graph. **Figure 11-111** shows a weighted, directed graph.

Figure 11-111 Weighted, directed graph



Based on the graph, each Router uses an SPF algorithm to calculate an SPT with itself as the root. The SPT shows routes to nodes in the AS. **Figure 11-112** shows an SPT.

Figure 11-112 SPT



When a Router's LSDB changes, the Router recalculates a shortest path. Frequent SPF calculations consume a large amount of resources and affect Router efficiency. Changing the interval between SPF calculations can prevent resource consumption caused by frequent LSDB changes. The default interval between SPF calculations is 5 seconds.

 Copyright © Huawei Technologies Co., Ltd.

The route calculation process is as follows:

1. A Router calculates intra-area routes.

   The Router uses an SFP algorithm to calculate shortest paths to other Routers in an area. Type 1 and Type 2 LSAs accurately describe the network topology in an area. Based on the network topology described by a Type 1 LSA, the Router calculates paths to other Routers in the area.

   NOTE

   If multiple equal-cost routes are produced during route calculation, the SPF algorithm retains all these routes in the LSDB.

2. The Router calculates inter-area routes.

   The network segment of the routes in an adjacent area can be considered to be directly connected to the area border router (ABR). Because the shortest path to the ABR has been calculated in the preceding phase, the Router can directly check a Type 3 LSA to obtain the shortest path to the network segment. The autonomous system boundary router (ASBR) can also be considered to be connected to the ABR. Therefore, the shortest path to the ASBR can also be calculated in this phase.

   NOTE

   If the Router performing an SPF calculation is an ABR, the Router needs to check only Type 3 LSAs in the backbone area.

3. The Router calculates AS external routes.

   AS external routes can be considered to be directly connected to the ASBR. Because the shortest path to the ASBR has been calculated in the preceding phase, the Router can check Type 5 LSAs to obtain the shortest paths to other ASs.

11.12.6.2.3 OSPF Route Control

You can use the following features to control the advertising and receiving of Open Shortest Path First (OSPF) routing information to import routes from other protocols:

- **Route summarization**

  Route summarization enables a Router to summarize routes with the same prefix into a single route and to advertise only the summarized route to other areas. Route summarization reduces the size of a routing table and improves Router performance.

- **Route filtering**

  OSPF can use routing policies to filter routes. By default, OSPF does not filter routes.

- **Smart-discover**

  Smart-discover enables a Router to send Hello packets to its neighbors without waiting for the Hello timer to expire, which speeds up neighbor relationship establishment.

Route Summarization

When a large OSPF network is deployed, an OSPF routing table includes a large number of routing entries. To accelerate route searches and simplify management,

 Copyright © Huawei Technologies Co., Ltd.

configure route summarization to reduce the size of the OSPF routing table. If a link frequently alternates between Up and Down, the links not involved in the route summarization are not affected. This process prevents route flapping and improves network stability.

Route summarization can be carried out in the following manners:

- Area border router (ABR) summarization

  When an ABR transmits routing information to other areas, it generates Type 3 LSAs for each network segment. If consecutive network segments exist in this area, you can manually summarize these network segments into a single network segment. The ABR generates one LSA for the summarized network segment and advertises only that LSA.

- Autonomous system boundary router (ASBR) summarization

  If route summarization has been configured and the local Router is an ASBR, the local Router summarizes imported Type 5 LSAs within the summarized address range. If a not-so-stubby area (NSSA) has been configured, the local Router also summarizes imported Type 7 LSAs within the summarized address range.

  If the local Router is both an ASBR and an ABR, it summarizes Type 5 LSAs translated from Type 7 LSAs.

Route Filtering

OSPF routing policies include access control lists (ACLs), IP prefix lists, and route-policies. For details about these policies, see the section "Routing Policy" in the *Feature Description - IP Routing*.

OSPF route filtering applies in the following aspects:

- Route import

  OSPF can import the routes learned by other routing protocols. A Router uses a configured routing policy to filter routes and imports only the routes matching the routing policy. Only an ASBR can import routes, and therefore a routing policy for importing routes must be configured on the ASBR.

- Advertising of imported routes

  A Router advertises imported routes to its neighbors. Only an ASBR can import routes, and therefore a routing policy for the advertising of imported routes must be configured on the ASBR.

- Route learning

  A Router uses a routing policy to filter received intra-area, inter-area, and AS external routes. The Router adds only the routes matching the routing policy to its routing table. All routes can still be advertised from an OSPF routing table.

  The Router filters only routes calculated based on LSAs, and therefore learned LSAs are complete.

- Inter-area LSA learning

  An ABR in an area can be configured to filter Type 3 LSAs advertised to the area. The ABR can advertise only Type 3 LSAs, and therefore a routing policy for inter-area LSA learning must be configured on the ABR.

 Copyright © Huawei Technologies Co., Ltd.

During inter-area LSA learning, the ABR directly filters Type 3 LSAs advertised to the area.

- Inter-area LSA advertising

  An ABR in an area can be configured to filter Type 3 LSAs advertised to other areas. The ABR can advertise only Type 3 LSAs, and therefore a routing policy for inter-area LSA advertising must be configured on the ABR.

Smart-Discover

Generally, Routers periodically send Hello packets through OSPF interfaces. OSPF uses a Hello timer to control the interval at which Hello packets are sent. A Router can send Hello packets again only when the Hello timer expires. Neighbors keep waiting to receive Hello packets until the Hello timer expires. This process delays the establishment of OSPF neighbor relationships. You can enable smart-discover to resolve this issue.

After you enable smart-discover on a Router, the Router sends Hello packets to its neighbors without waiting for the Hello timer to expire. Smart-discover ensures that neighbors can quickly receive Hello packets and perform a fast state transition.

Enable smart-discover when either of the following cases is true:

- The neighbor status becomes 2-way for the first time.
- The neighbor status changes from 2-way or higher to Init.

11.12.6.2.4 OSPF NSSA

Background

As defined in the Open Shortest Path First (OSPF) protocol, stub areas cannot import external routes. This mechanism prevents external routes from consuming the bandwidth and storage resources of Routers in stub areas. If you need to both import external routes and prevent resource consumption caused by external routes, you can configure not-so-stubby areas (NSSAs).

Derived from stub areas, NSSAs resemble stub areas in many ways. Different from stub areas, NSSAs can import autonomous system (AS) external routes and advertise them within the entire AS, without learning external routes from other areas in the AS.

Related Concepts

- **N-bit**

  A Router uses the N-bit carried in a Hello packet to identify the area type that it supports. The same area type must be configured for all Routers in an area. If Routers have different area types, they cannot establish OSPF neighbor relationships. Some vendors' devices do not comply with RFC 1587, and the N-bit is also set in OSPF Database Description (DD) packets. You can manually set the N-bit on a Router to interwork with the vendors' devices.

- **Type 7 LSA**

  Type 7 LSAs, which describe imported external routes, are introduced to support NSSAs. Type 7 LSAs are generated by an autonomous system

 Copyright © Huawei Technologies Co., Ltd.

boundary router (ASBR) in an NSSA and advertised only within the NSSA. After an area border router (ABR) in an NSSA receives Type 7 LSAs, it selectively translates Type 7 LSAs into Type 5 LSAs to advertise external routes to other areas on an OSPF network.

Principles

To advertise external routes imported by an NSSA to other areas, a translator must translate Type 7 LSAs into Type 5 LSAs. Notes for an NSSA are as follows:

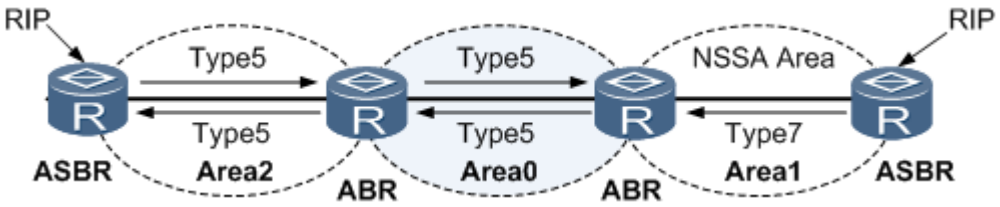
- By default, the translator is the ABR with the largest router ID in the NSSA.
- The propagate bit (P-bit) is used to notify a translator whether Type 7 LSAs need to be translated.
- Only Type 7 LSAs with the P-bit set and a non-zero forwarding address (FA) can be translated into Type 5 LSAs. An FA indicates that packets to a destination address will be forwarded to the address specified by the FA.

  NOTE

  FA indicates that the packet to a specific destination address is to be forwarded to the address specified by.

  The loopback interface address in an area is preferentially selected as the FA. If no loopback interface exists, the address of the interface that is Up and has the largest logical index in the area is selected as the FA.
- The P-bit is not set for Type 7 LSAs generated by an ABR.

Figure 11-113 NSSA



Advantages

Multiple ABRs may be deployed in an NSSA. To prevent routing loops caused by default routes, ABRs do not calculate the default routes advertised by each other.

11.12.6.2.5 OSPF Fast Convergence

OSPF fast convergence is an extended feature of OSPF implemented to speed up the convergence of routes. It includes the following components:

- Incremental SPF (I-SPF): recalculates only the routes of the changed nodes rather than all the nodes when the network topology changes. This speeds up route calculation.
- Partial Route Calculation (PRC): calculates only the changed routes when the routes on the network change.
- An OSPF intelligent timer: can dynamically adjust its value based on the user's configuration and the interval at which an event is triggered, such as the route calculation interval, which ensures rapid and stable network operation.

 Copyright © Huawei Technologies Co., Ltd.

OSPF intelligent timer applies the exponential backoff technology so that the value of the timer can reach the millisecond level.

I-SPF

In ISO 10589, the Dijkstra algorithm was adopted to calculate routes. When a node changes on the network, this algorithm is used to recalculate all routes. The calculation takes a long time and consumes too many CPU resources, which affects the convergence speed.

I-SPF improves the Dijkstra algorithm. Except for the first time, only changed nodes instead of all nodes are involved in calculation. The SPT generated at last is the same as that generated by the previous algorithm, but I-SPF decreases CPU usage and speeds up network convergence.

PRC

Similar to I-SPF, PRC calculates only the changed routes. However, PRC does not calculate the shortest path. It updates routes based on the SPT calculated by I-SPF.

In route calculation, a leaf represents a route, and a node represents a router. SPT and leaf changes both cause the change of routing information, but the SPT change is irrelevant to the leaf change. PRC processes routing information based on either SPT or leaf information.

- If the SPT changes, PRC processes the routing information of all leaves on a changed node.
- If the SPT does not change, PRC does not process the routing information on any node.
- If a leaf changes, RPC processes the routing information on the leaf only.
- If a leaf does not change, PRC does not process the routing information on any leaf.

For example, if OSPF is enabled on an interface of a node, the SPT calculated by I-SPF remains unchanged. PRC updates only the routes of this interface, consuming less CPU resources.

PRC improves on the SPF algorithm. Working with I-SPF, RPC further improves the convergence performance of the network.

NOTE

In the implementation of a device, only I-SPF and PRC are used to calculate OSPF routes.

OSPF Intelligent Timer

On an unstable network, routes are calculated frequently, which consumes a great number of CPU resources. In addition, LSPs that describe the unstable topology are generated and transmitted on the unstable network. Frequently processing such LSAs affects the rapid and stable operation of the entire network.

To speed up route convergence on the entire network, the OSPF intelligent timer controls route calculation, LSA generation, and LSA receiving.

OSPF intelligent timer speeds up route convergence in the following modes:

 Copyright © Huawei Technologies Co., Ltd.

- On a network where routes are calculated repeatedly, the OSPF intelligent timer dynamically adjusts the route calculation based on user's configuration and the exponential backoff technology. The number of route calculation times and the CPU resource consumption are decreased. Routes are calculated after the network topology becomes stable.
- On an unstable network, if a router generates or receives LSAs due to frequent topology changes, the OSPF intelligent timer can dynamically adjust its value. No LSA is generated or handled within an interval, which prevents invalid LSAs from being generated and advertised on the entire network.

The OSPF intelligent timer is started by default and uses the default value.

11.12.6.2.6 OSPF NSR

Non-Stop Routing (NSR) is a routing technique that prevents neighboring routers from detecting faults on the control plane of a device with a backup control plane. With NSR, when the control plane of the device fails, the neighbor relationship set up through specific routing protocols, MPLS, and other protocols that carry services continue uninterrupted.

As networks develop and expand, carriers place increasingly higher requirements on the reliability of IP networks. NSR was introduced as a high availability (HA) solution to ensure that services transmitted by a device were unaffected by hardware or software failures on the device.

OSPF NSR synchronizes the protocol data on the master main control board to the slave main control board in real time. When the master main control board fails or needs to be upgraded, the slave main control board immediately takes over services from the master main control board, and the neighbor remains unaware the fault occurred. Real-time data synchronization between the master and slave main control boards through OSPF NSR is implemented as follows:

- OSPF backs up configuration data and dynamic data, including information about interfaces, neighbors, and LSDBs.
- OSPF does not back up routes, shortest path trees (SPTs), or Traffic Engineering Databases (TEDBs). Such information can be restored using the source data in the database backup process.
- When the master/slave main control board switchover occurs, the new master main control board restores the operations data and takes over services from the former master main control board. Throughout this process, the neighbor remains unaware of the fault.

11.12.6.2.7 OSPF Authentication

OSPF authentication encrypts OSPF packets by adding the authentication field to packets to ensure network security. When a local device receives OSPF packets from a remote device, the local device discards the packets if the authentication passwords carried in these packets do not match the local one, which protects the local device from potential attacks.

In terms of the packet type, the authentication is classified as follows:

- Area authentication

  Area authentication is configured in the OSPF area view and applies to packets received by all interfaces in the OSPF area.

 Copyright © Huawei Technologies Co., Ltd.

- Interface authentication

  Interface authentication is configured in the interface view and applies to all packets received by the interface.

In terms of packet the authentication modes, the authentication is classified as follows:

- Non-authentication

  Authentication is not required.
- Simple authentication

  The authenticated party directly adds the configured password to packets for authentication. This authentication mode provides the lowest password security.
- MD5 authentication

  The authenticated party encrypts the configured password using a Message Digest 5 (MD5) algorithm and adds the ciphertext password to packets for authentication. This authentication mode improves password security. The supported MD5 algorithms include MD5 and HMAC-MD5.

  NOTICE

  MD5 and HMAC-MD5 are insecure encryption algorithms. Therefore, using HMAC-SHA256 is recommended.
- HMAC-SHA256 authentication

  A password is encrypted using the HMAC-SHA256 algorithm before it is added to the packet, which improves password security.

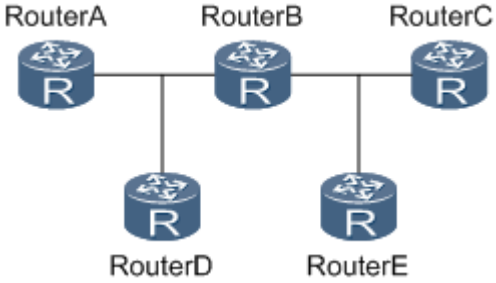
OSPF carries authentication types in packet headers and authentication information in packet tails.

The authentication types include:

- 0: Non-authentication
- 1: Simple authentication
- 2: Ciphertext authentication

Usage Scenario

Figure 11-114 Networking for OSPF authentication on a broadcast network



The configuration requirements are as follows:

 Copyright © Huawei Technologies Co., Ltd.

- The interface authentication configurations must be the same on all devices on the same network so that OSPF neighbor relationships can be established.
- The area authentication configurations must be the same on all devices in the same area.

11.12.6.2.8 OSPF Packet Format

Open Shortest Path First (OSPF) packets are encapsulated into IP packets. The OSPF protocol number is 89. OSPF packets are classified into the following types:

- **Hello packet**
- **Database Description (DD) packet**
- **Link State Request (LSR) packet**
- **Link State Update (LSU) packet**
- **Link State Acknowledgment (LSAck) packet**

Packet Header Format

The five types of OSPF packets have the same packet header format. The length of an OSPF packet header is 24 bytes. **Figure 11-115** shows an OSPF packet header.

Figure 11-115 OSPF packet header

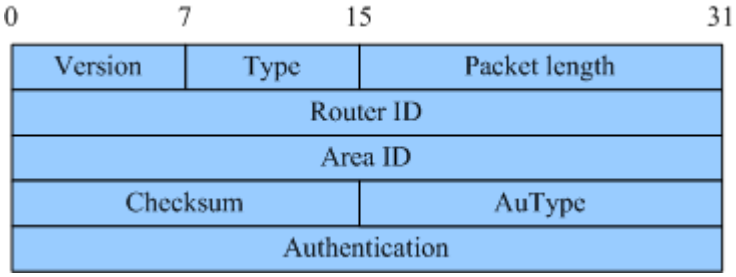


Table 11-42 Packet header fields

Field	Length	Description
Version	8 bits	OSPF version number. For OSPFv2, the value is 2.
Type	8 bits	OSPF packet type. The values are as follows: • 1: Hello packet • 2: DD packet • 3: LSR packet • 4: LSU packet • 5: LSAck packet
Packet length	16 bits	Length of the OSPF packet containing the packet header, in bytes.
Router ID	32 bits	ID of the Router that sends the OSPF packet.

 Copyright © Huawei Technologies Co., Ltd.

Field	Length	Description
Area ID	32 bits	ID of the area to which the Router that sends the OSPF packet belongs.
Checksum	16 bits	Checksum of the OSPF packet that does not contain the Authentication field.
AuType	16 bits	Authentication type. The values are as follows: • 0: non-authentication • 1: simple authentication • 2: message digest algorithm 5 (MD5) authentication **NOTICE** MD5 is an insecure encryption algorithm. Therefore, using HMAC-SHA256 is recommended.
Authentication	64 bits	This field has different meanings for different AuType values: • 0: This field is not defined. • 1: This field defines password information. • 2: This field contains the key ID, MD5 authentication data length, and sequence number.

NOTE

MD5 authentication data is added to an OSPF packet and is not included in the Authentication field.

Hello Packet

Hello packets are commonly used packets, which are periodically sent on OSPF interfaces to establish and maintain neighbor relationships. A Hello packet includes information about the designated router (DR), backup designated router (BDR), timers, and known neighbors. **Figure 11-116** shows the format of a Hello packet.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-116 Format of a Hello packet

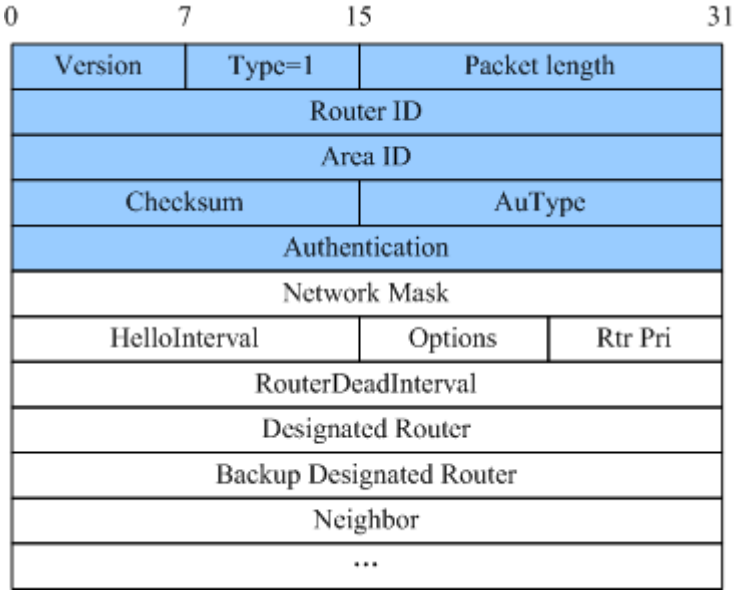


Table 11-43 Hello packet fields

Field	Length	Description
Network Mask	32 bits	Mask of the network on which the interface that sends the Hello packet resides.
HelloInterval	16 bits	Interval at which Hello packets are sent.
Options	8 bits	The values are as follows: • E: Type 5 link state advertisements (LSAs) are flooded. • MC: IP multicast packets are forwarded. • N/P: Type 7 LSAs are processed. • DC: On-demand links are processed.
Rtr Pri	8 bits	DR priority. The default value is 1. **NOTE** If the DR priority of a Router interface is set to 0, the interface cannot participate in a DR or BDR election.
RouterDeadInterval	32 bits	Dead interval. If a Router does not receive any Hello packets from its neighbors within a specified dead interval, the neighbors are considered Down.
Designated Router	32 bits	Interface address of the DR.
Backup Designated Router	32 bits	Interface address of the BDR.
Neighbor	32 bits	Router ID of the neighbor.

 Copyright © Huawei Technologies Co., Ltd.

Table 11-44 lists the address types, interval types, and default intervals used when Hello packets are transmitted on different networks.

Table 11-44 Hello packet characteristics for various network types

Network Type	Address Type	Interval Type	Default Interval
Broadcast	Multicast address	HelloInterval	10 seconds
Non-broadcast multiple access (NBMA)	Unicast address	• HelloInterval for the DR, BDR, and Router that can become a DR • PollInterval for the case when neighbors become Down and HelloInterval for other cases	30 seconds for HelloInterval 120 seconds for PollInterval
Point-to-point (P2P)	Multicast address	HelloInterval	10 seconds
Point-to-multipoint (P2MP)	Unicast address	HelloInterval	30 seconds

NOTE

To establish neighbor relationships between Routers on the same network segment, you must set the same HelloInterval, PollInterval, and RouterDeadInterval values for the Routers. PollInterval applies only to NBMA networks.

DD Packet

During an adjacency initialization, two Routers use DD packets to describe their own link state databases (LSDBs) for LSDB synchronization. A DD packet contains the header of each LSA in an LSDB. An LSA header uniquely identifies an LSA. The LSA header occupies only a small portion of the LSA, which reduces the amount of traffic transmitted between Routers. A neighbor can use the LSA header to check whether it already has the LSA. When two Routers exchange DD packets, one functions as the master and the other functions as the slave. The master defines a start sequence number. The master increases the sequence number by one each time it sends a DD packet. After the slave receives a DD packet, it uses the sequence number carried in the DD packet for acknowledgement.

Figure 11-117 shows the format of a DD packet.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-117 Format of a DD packet

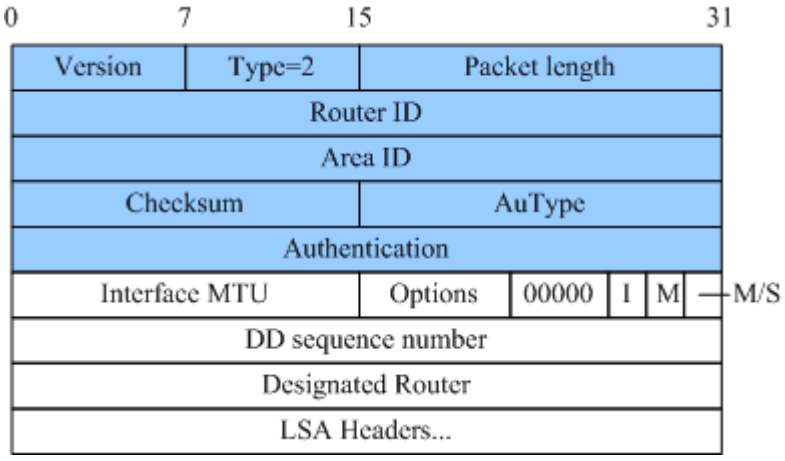


Table 11-45 DD packet fields

Field	Length	Description
Interface MTU	16 bits	Maximum length of the DD packet sent by the interface with packet fragmentation disabled.
Options	8 bits	The values are as follows: • E: Type 5 LSAs are flooded. • MC: IP multicast packets are forwarded. • N/P: Type 7 LSAs are processed. • DC: On-demand links are processed.
I	1 bit	If the DD packet is the first packet among multiple consecutive DD packets sent by a Router, this field is set to 1. In other cases, this field is set to 0.
M (More)	1 bit	If the DD packet is the last packet among multiple consecutive DD packets sent by a Router, this field is set to 0. In other cases, this field is set to 1.
M/S (Master/ Slave)	1 bit	When two Routers exchange DD packets, they negotiate a master/slave relationship. The Router with a larger router ID becomes the master. If this field is set to 1, the DD packet is sent by the master.
DD sequence number	32 bits	Sequence number of the DD packet. The master and slave use the sequence number to ensure that DD packets are correctly transmitted.
LSA Headers	-	LSA header information included in the DD packet.

 Copyright © Huawei Technologies Co., Ltd.

LSR Packet

After two Routers exchange DD packets, they send LSR packets to request each other's LSAs. The LSR packets contain the summaries of the requested LSAs. **Figure 11-118** shows the format of an LSR packet.

Figure 11-118 Format of an LSR packet

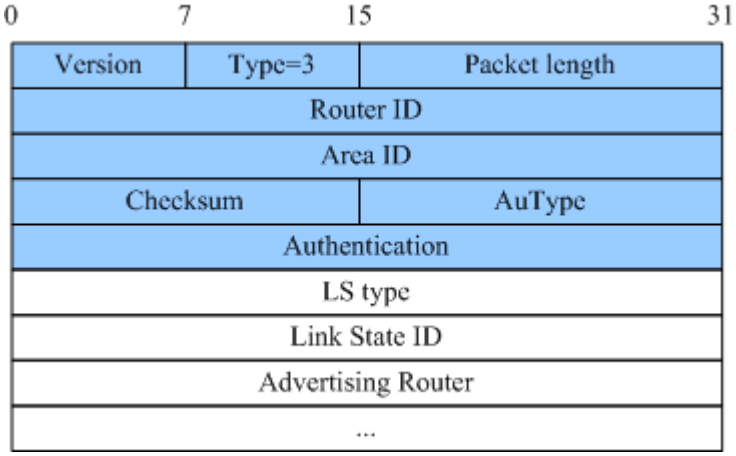


Table 11-46 LSR packet fields

Field	Length	Description
LS type	32 bits	Type of the LSA
Link State ID	32 bits	This field together with the LS type field describes an LSA in an AS.
Advertising Router	32 bits	Router ID of the Router that generates the LSA.

NOTE

The LS type, Link State ID, and Advertising Router fields can uniquely identify an LSA. If two LSAs have the same LS type, Link State ID, and Advertising Router fields, a Router uses the LS sequence number, LS checksum, and LS age fields to obtain a required LSA.

LSU Packet

A Router uses an LSU packet to transmit LSAs requested by its neighbors or to flood its own updated LSAs. The LSU packet contains a set of LSAs. For multicast and broadcast networks, LSU packets are multicast to flood LSAs. To ensure reliable LSA flooding, a Router uses an LSAck packet to acknowledge the LSAs contained in an LSU packet that is received from a neighbor. If an LSA fails to be acknowledged, the Router retransmits the LSA to the neighbor. **Figure 11-119** shows the format of an LSU packet.

Copyright © Huawei Technologies Co., Ltd.

Figure 11-119 Format of an LSU packet

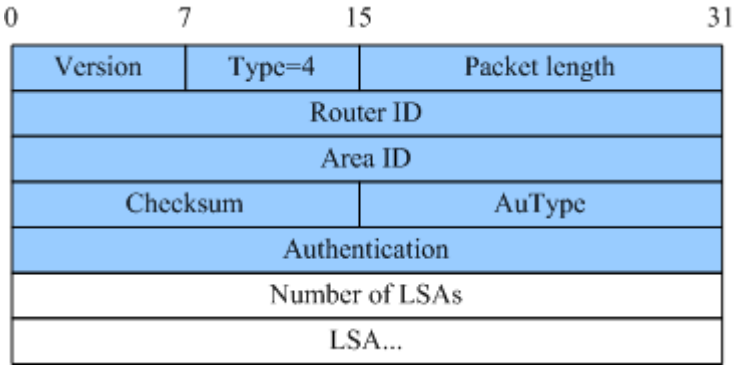


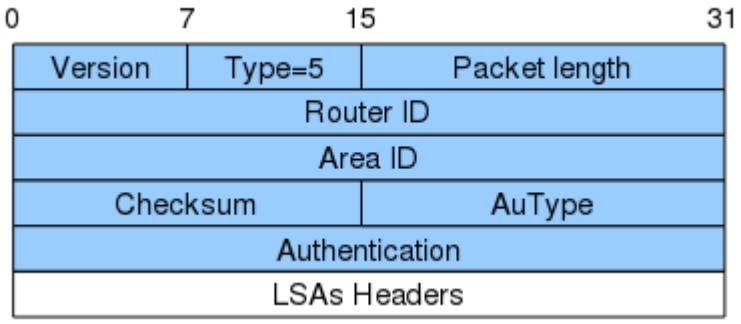
Table 11-47 LSU packet field

Field	Length	Description
Number of LSAs	32 bits	Number of LSAs contained in the LSU packet

LSAck Packet

A Router uses an LSAck packet to acknowledge the LSAs contained in a received LSU packet. The LSAs can be acknowledged using LSA headers. LSAck packets can be transmitted over different links in unicast or multicast mode. **Figure 11-120** shows the format of an LSAck packet.

Figure 11-120 Format of an LSAck packet



 Copyright © Huawei Technologies Co., Ltd.

Table 11-48 LSAck packet field

Field	Length	Description
LSAs Headers	Determined by the header length of the LSA to be acknowledged.	This field is used to acknowledge an LSA.

11.12.6.2.9 OSPF LSA Format

Each Router in an autonomous system (AS) generates one or more types of link state advertisements (LSAs), depending on the Router's type. Multiple LSAs form a link state database (LSDB). Open Shortest Path First (OSPF) encapsulates routing information into LSAs for transmission. Commonly used LSAs include:

- **Router-LSAs (Type 1)**
- **Network-LSAs (Type 2)**
- **Summary-LSAs**, including network-summary-LSAs (Type 3) and ASBR-summary-LSAs (Type 4)
- **AS-external-LSAs (Type 5)**

LSA Header Format

All LSAs have the same header. **Figure 11-121** shows an LSA header.

Figure 11-121 LSA header

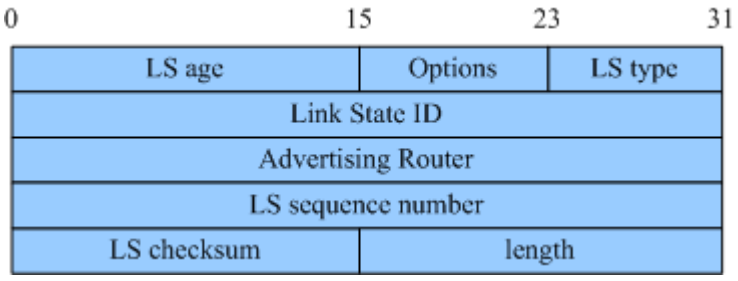


Table 11-49 LSA header fields

Field	Length	Description
LS age	16 bits	Time that elapses after the LSA is generated, in seconds. The value of this field continually increases regardless of whether the LSA is transmitted over a link or saved in an LSDB.

 Copyright © Huawei Technologies Co., Ltd.

Field	Length	Description
Options	8 bits	The values are as follows: • E: Type 5 LSAs are flooded. • MC: IP multicast packets are forwarded. • N/P: Type 7 LSAs are processed. • DC: On-demand links are processed.
LS type	8 bits	Type of the LSA. The values are as follows: • Type1: Router-LSA • Type2: Network-LSA • Type3: Network-summary-LSA • Type4: ASBR-summary-LSA • Type5: AS-external-LSA • Type7: NSSA-LSA
Link State ID	32 bits	This field together with the LS type field describes an LSA in an AS.
Advertising Router	32 bits	Router ID of the Router that generates the LSA.
LS sequence number	32 bits	Sequence number of the LSA. Neighbors can use this field to identify the latest LSA.
LS checksum	16 bits	Checksum of all fields except the LS age field.
length	16 bits	Length of the LSA including the LSA header, in bytes.

Router-LSA

A router-LSA describes the link status and cost of a Router. Router-LSAs are generated by a Router and advertised within the area to which the Router belongs. **Figure 11-122** shows the format of a router-LSA.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-122 Format of a router-LSA

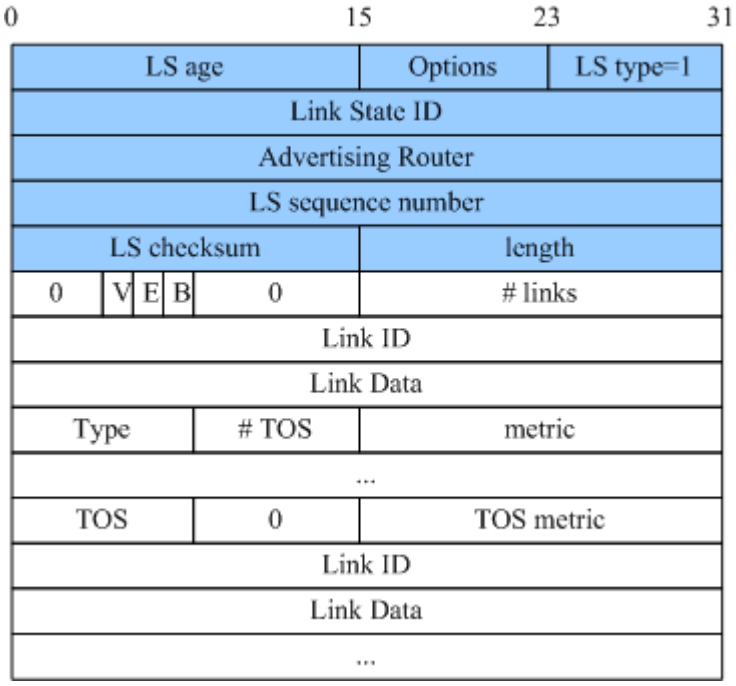


Table 11-50 Router-LSA fields

Field	Length	Description
Link State ID	32 bits	Router ID of the Router that generates the LSA.
V (Virtual Link)	1 bit	If the Router that generates the LSA is located at one end of a virtual link, this field is set to 1. In other cases, this field is set to 0.
E (External)	1 bit	If the Router that generates the LSA is an autonomous system boundary router (ASBR), this field is set to 1. In other cases, this field is set to 0.
B (Border)	1 bit	If the Router that generates the LSA is an area border router (ABR), this field is set to 1. In other cases, this field is set to 0.
# links	16 bits	Number of links and interfaces described in the LSA, including all links and interfaces in the area to which the Router belongs.
Link ID	32 bits	Object to which the Router is connected. This field has different meanings for different link types: ● 1: router ID ● 2: interface IP address of the designated router (DR) ● 3: network segment or subnet number ● 4: router ID of the neighbor on a virtual link

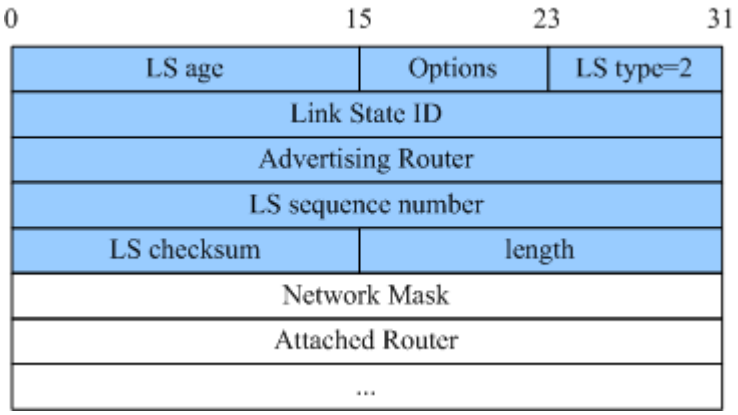
 Copyright © Huawei Technologies Co., Ltd.

Field	Length	Description
Link Data	32 bits	Link data. This field has different meanings for different link types: • 1: interface index • 3: subnet mask • 2 and 4: interface address of the Router
Type	8 bits	Type of the Router link. The values are as follows: • 1: The Router is connected to another Router in point-to-point (P2P) mode. • 2: The Router is connected to a transport network. • 3: The Router is connected to a stub network. • 4: The Router is connected to another Router over a virtual link.
# TOS	8 bits	Number of types of service (ToSs).
metric	16 bits	Cost of the link.
TOS	8 bits	Type of service.
TOS metric	16 bits	Metric for the specified ToS.

Network-LSA

A network-LSA describes the link status of all Routers on the local network segment. Network-LSAs are generated by a DR on a broadcast or non-broadcast multiple access (NBMA) network and advertised within the area to which the DR belongs. **Figure 11-123** shows the format of a network-LSA.

Figure 11-123 Format of a network-LSA



 Copyright © Huawei Technologies Co., Ltd.

Table 11-51 Network-LSA fields

Field	Length	Description
Link State ID	32 bits	Interface IP address of the DR
Network Mask	32 bits	Mask of the broadcast or NBMA network
Attached Router	32 bits	Router IDs of all Routers on the broadcast or NBMA network, including the router ID of the DR

Summary-LSA

A network-summary-LSA describes routes on a network segment in an area. The routes are advertised to other areas.

An ASBR-summary-LSA describes routes to the ASBR in an area. The routes are advertised to all areas except the area to which the ASBR belongs.

The two types of summary-LSAs have the same format and are generated by an ABR. **Figure 11-124** shows the format of a summary-LSA.

Figure 11-124 Format of a summary-LSA

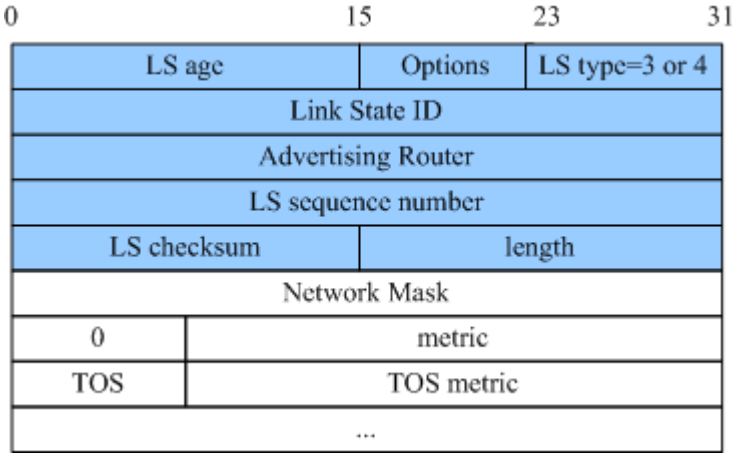


Table 11-52 Network-summary-LSA fields

Field	Length	Description
Link State ID	32 bits	Advertised network address
Network Mask	32 bits	Mask of the broadcast or NBMA network
metric	24 bits	Cost of the route to the destination address
TOS	8 bits	Type of service.

Copyright © Huawei Technologies Co., Ltd.

Field	Length	Description
TOS metric	24 bits	Metric for the specified ToS

NOTE

When a default route is advertised, both the Link State ID and Network Mask fields are set to 0.0.0.0.

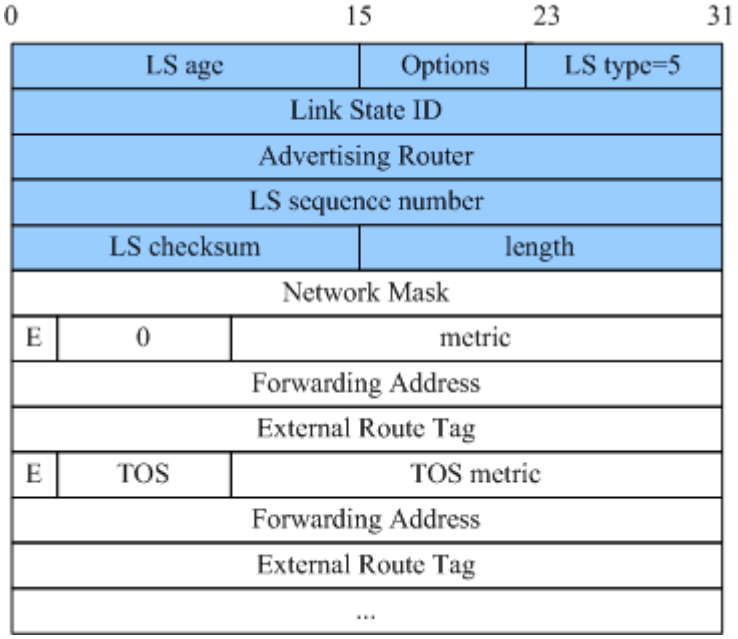
Table 11-53 ASBR-summary-LSA fields

Field	Length	Description
Link State ID	32 bits	Router ID of the ASBR
Network Mask	32 bits	Set to 0.0.0.0
metric	24 bits	Cost of the route to the destination address
TOS	8 bits	Type of service.
TOS metric	24 bits	Metric for the specified ToS

AS-External-LSA

An AS-external-LSA describes AS external routes. AS-external-LSAs are generated by an ASBR. Among the five types of LSAs, only AS-external-LSAs can be advertised to all areas except stub areas and not-so-stubby areas (NSSAs). **Figure 11-125** shows the format of an AS-external-LSA.

Figure 11-125 Format of an AS-external-LSA



 Copyright © Huawei Technologies Co., Ltd.

Table 11-54 AS-external-LSA fields

Field	Length	Description
Link State ID	32 bits	Advertised network address
Network Mask	32 bits	Mask of the advertised destination address.
E	1 bit	Type of the external route. The values are as follows: ● 0: Type 1 external route ● 1: Type 2 external route
metric	24 bits	Cost of the route to the destination address
Forwarding Address	32 bits	Packets destined for the advertised destination address are forwarded to the address specified by this field.
External Route Tag	32 bits	Tag added to the external route. This field can be used to manage external routes. OSPF itself does not use this field.
TOS	8 bits	Type of service.
TOS metric	24 bits	Metric for the specified ToS.

NOTE

When AS-external-LSAs are used to advertise default routes, both the Link State ID and Network Mask fields are set to 0.0.0.0.

11.12.6.2.10 OSPF IP FRR

OSPF IP fast reroute (FRR) is a kind of dynamic IP FRR. To implement this feature, Interior Gateway Protocol (IGP) calculates the backup paths beforehand by using the link status database of the entire network and then stores the backup paths in the forwarding table to minimize service interruption duration in case of a link fault.

Background

As networks develop, voice over IP (VoIP) and online video services pose higher requirements for real-time transmission. Nevertheless, if a primary link fails, OSPF-enabled devices need to perform multiple operations, including detecting the fault, updating the link-state advertisement (LSA), flooding the LSA, calculating routes, and delivering forward information base (FIB) entries before switching traffic to a new link. This process takes a time much longer than the minimum delay that can cause users to sense service interruption. As a result, the requirements for real-time transmission cannot be met. OSPF IP FRR can solve this problem. OSPF IP FRR conforms to dynamic IP FRR defined by RFC 5286. With OSPF IP FRR, devices can quickly switch traffic from a faulty primary link to a backup link, protecting against a link or node failure.

 Copyright © Huawei Technologies Co., Ltd.

Related Concepts

OSPF IP FRR

OSPF IP FRR refers to a mechanism in which a device uses the loop-free alternate (LFA) algorithm to compute the next hop of a backup link and stores the next hop together with the primary link in the forwarding table. If the primary link fails, the device switches the traffic to the backup link before routes are converged on the control plane. This mechanism reduces the traffic interruption duration and minimizes the impacts.

OSPF IP FRR Policy

An OSPF IP FRR policy can be configured to filter alternate next hops. Only the alternate next hops that match the filtering rules of the policy can be added to the IP routing table.

LFA algorithm

A device uses each neighbor that can be used as a backup next hop as a root node, uses the shortest path first (SPF) algorithm to calculate the shortest path from each root node to the destination, and then uses the inequalities defined in RFC 5286 to calculate the backup link that is loop-free and has the lowest cost.

Networking

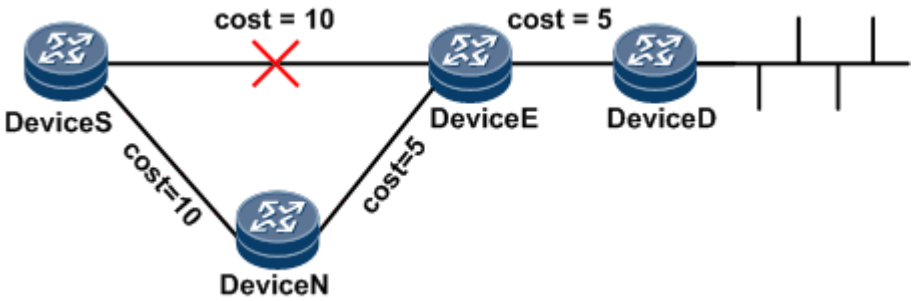
OSPF IP FRR protects traffic from a fault on a specified link.

In **Figure 11-126**, traffic flows from Device S to Device D. The primary link is Device S->Device E->Device D, and the backup link is Device S->Device N->Device E->Device D. If link costs meet the inequality: Distance_opt(N, D) < Distance_opt(N, S) + Distance_opt(S, D) and OSPF IP FRR is enabled, Device S switches the traffic to the backup link if the primary link fails, reducing the traffic interruption duration.

NOTE

Distance_opt(X, Y) indicates the shortest link from X to Y. S stands for a source node, E for the faulty node, N for a node along a backup link, and D for a destination node.

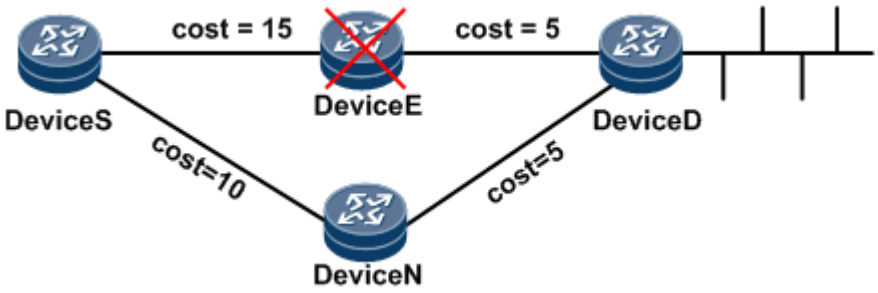
Figure 11-126 OSPF IP FRR link protection



The switching process in the following figure is similar. Therefore, the details are omitted.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-127 OSPF IP FRR node-and-link protection



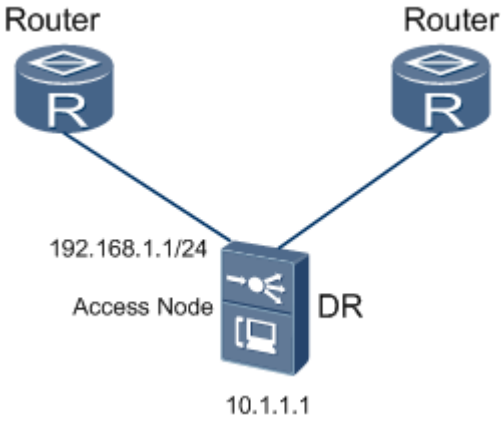
11.12.6.3 Configuration Example of OSPF

This topic provides an example for configuring OSPF on the MA5800.

Service Requirements

- OSPF is enabled on the MA5800.

Figure 11-128 Example network for configuring OSPF



Data Plan

Table 11-55 provides the data plan for configuring OSPF.

Table 11-55 Data plan for configuring OSPF

Item	Data
MA5800	IP address of the Layer 3 interface: 192.168.1.1/24
	Priority: 100
	VLAN ID: 2
	Router ID: 10.1.1.1

 Copyright © Huawei Technologies Co., Ltd.

Prerequisite

- The native VLAN of each interface of the MA5800 must be configured to ensure a normal communication.
- The OSPF area IDs of the MA5800 device and the routers must be consistent.

Procedure

Step 1 Configure MA5800.

1. Configure the IP address of the Layer 3 interface.

```
huawei(config)#vlan 2 smart
huawei(config)#port vlan 2 0/9 0
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 192.168.1.1 24
huawei(config-if-vlanif2)#quit
```

2. Configure the OSPF Router ID.

```
huawei(config)#router id 10.1.1.1
```

3. Enable OSPF.

```
huawei(config)#ospf
huawei(config-ospf-1)#area 0
huawei(config-ospf-1-area-0.0.0.0)#network 192.168.1.0 0.0.0.255
huawei(config-ospf-1-area-0.0.0.0)#network 10.1.1.1 0.0.0.0
huawei(config-ospf-1-area-0.0.0.0)#quit
huawei(config-ospf-1)#quit
```

4. Configure the OSPF priority.

```
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ospf dr-priority 100
huawei(config-if-vlanif2)#quit
```

5. Save the data.

```
huawei(config)#save
```

----End

Result

Run the **display ip routing-table** command and you can find the learned route table. Hosts can communicate with each other.

Configuration File

```
vlan 2 smart
port vlan 2 0/9 0
interface vlanif 2
ip address 192.168.1.1 24
quit
router id 10.1.1.1
ospf
area 0
network 192.168.1.0 0.0.0.255
network 10.1.1.1 0.0.0.0
quit
quit
interface vlanif 2
ospf dr-priority 100
quit
save
```

 Copyright © Huawei Technologies Co., Ltd.

11.12.6.4 Example for Configuring OSPF IP FRR

OSPF IP FRR can generate a loop-free backup link for fast link switching in case of a link failure.

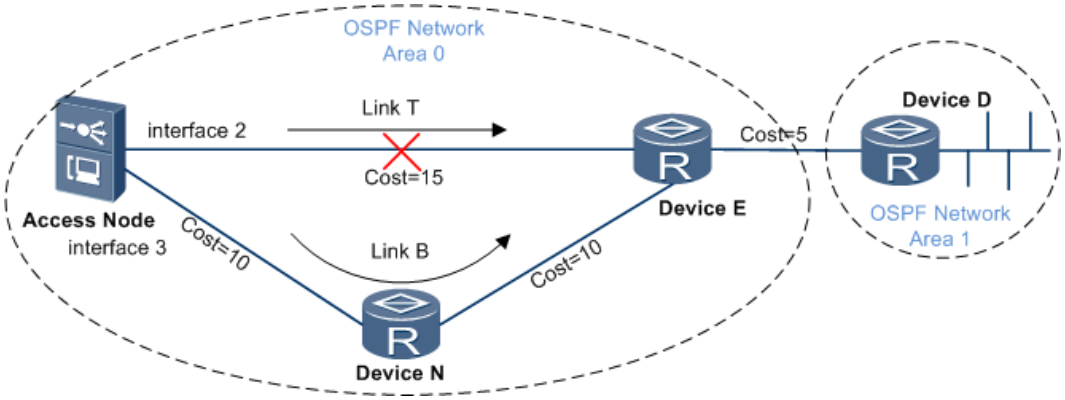
Service Requirements

Configuring OSPF IP FRR quickly switches traffic to the backup link without waiting for route convergence.

As shown in **Figure 11-129**,

- OSPF is running on all devices.
- The link cost meets the OSPF IP FRR traffic protection inequality.
- If the primary link T fails, traffic forwarded by the MA5800 is quickly switched to the backup link B.

Figure 11-129 Networking of configuring OSPF IP FRR



Data Plan

Table 11-56 describes data plan for configuring OSPF IP FRR.

Table 11-56 Data plan for configuring OSPF IP FRR

Item	Data
MA5800 port 2	IP address of the port: 10.1.2.1/24
	VLAN ID:3
	Cost:15
MA5800 port 3	IP address of the port: 10.1.3.1/24
	VLAN ID:4
	Cost:10

 Copyright © Huawei Technologies Co., Ltd.

Prerequisite

Before you configure OSPF IP FRR, complete the following tasks:

- Configuring link layer protocol parameters and IP addresses for ports to ensure that the link layer protocol on the ports is Up
- Enabling basic OSPF functions on routers
- Configuring basic BFD for OSPF functions on all devices in area 0 to quickly detect faults

Procedure

Step 1 Perform configuration on the MA5800.

1. Configure the IP address of the Layer 3 interface on the MA5800.

```
huawei(config)#vlan 3 smart
huawei(config)#port vlan 3 0/9 1
huawei(config)#interface vlanif 3
huawei(config-if-vlanif3)#ip address 10.1.2.1 24
huawei(config-if-vlanif3)#quit
huawei(config)#vlan 4 smart
huawei(config)#port vlan 4 0/9 2
huawei(config)#interface vlanif 4
huawei(config-if-vlanif4)#ip address 10.1.3.1 24
huawei(config-if-vlanif4)#quit
```

2. Configure the cost so that the traffic preferentially goes through link T.

```
huawei(config)#interface vlanif 3
huawei(config-if-vlanif3)#ospf cost 15
huawei(config-if-vlanif3)#quit
huawei(config)#interface vlanif 4
huawei(config-if-vlanif4)#ospf cost 10
huawei(config-if-vlanif4)#quit
```

3. Enable OSPF IP FRR on the MA5800.

```
huawei(config)#ospf
huawei(config-ospf-1)#frr
huawei(config-ospf-1-frr)#loop-free-alternate
huawei(config-ospf-1-frr)#quit
huawei(config-ospf-1)#quit
```

4. Save the data.

```
huawei(config)#save
```

----End

Result

Run the **display ospf routing** command to query the information about primary/secondary routes configured.

Configuration File

```
vlan 3 smart
port vlan 3 0/9 1
interface vlanif 3
ip address 10.1.2.1 24
quit
vlan 4 smart
port vlan 4 0/9 2
interface vlanif 4
```

 Copyright © Huawei Technologies Co., Ltd.

```
ip address 10.1.3.1 24
quit
interface vlanif 3
ospf cost 15
quit
interface vlanif 4
ospf cost 10
quit
ospf
frr
loop-free-alternate
quit
quit
save
```

11.12.6.5 OSPF Reference Standards and Protocols

Reference documents for OSPF are as follows:

Document	Description	Remarks
RFC 1583	OSPF Version 2	–
RFC 1765	OSPF Database Overflow	–
RFC 2328	This memo documents OSPFv2. Implementations of this memo are interoperable with RFC 1583.	–
RFC 2370	The OSPF Opaque LSA Option	–
RFC 3137	OSPF Stub Router Advertisement	This RFC is informational and non-standard.
RFC 3623	Graceful OSPF Restart	Partially supported by Huawei devices.
RFC 3630	Traffic Engineering (TE) Extensions to OSPF Version 2	–
RFC 3682	The Generalized TTL Security Mechanism (GTSM)	This RFC is experimental and non-standard.
RFC 3906	Calculating Interior Gateway Protocol (IGP) Routes Over Traffic Engineering Tunnels	-
RFC 4576	Using a Link State Advertisement (LSA) Options Bit to Prevent Looping in BGP/MPLS IP Virtual Private Networks (VPNs)	-
RFC 4577	PDF OSPF as the Provider/Customer Edge Protocol for BGP/MPLS IP Virtual Private Networks (VPNs)	-

 Copyright © Huawei Technologies Co., Ltd.

Document	Description	Remarks
RFC 4750	OSPF Version 2 Management Information Base	-

11.12.7 OSPFv3

11.12.7.1 Introduction

Definition

Open Shortest Path First (OSPF), developed by the Internet Engineering Task Force (IETF), is a link-state Interior Gateway Protocol (IGP).

At present, OSPF Version 2 (OSPFv2) is used for IPv4, while OSPF Version 3 (OSPFv3), developed on the basis of OSPFv2, is used for IPv6.

Purpose

The primary purpose of OSPFv3 is to develop a routing protocol independent of any specific network layer. The internal OSPFv3 router information is redesigned to achieve this purpose.

11.12.7.2 Principles

11.12.7.2.1 OSPFv3 Fundamentals

Running on IPv6, OSPFv3 (defined in RFC 2740) is an independent routing protocol that is developed on the basis of OSPFv2.

- OSPFv3 and OSPFv2 are the same in respect of the working principles of the Hello packet, state machine, link-state database (LSDB), flooding, and route calculation.
- OSPFv3 packets are encapsulated into IPv6 packets and can be transmitted in unicast or multicast mode.

OSPFv3 Packet Types

Packet Type	Function
Hello packet	Hello packets are sent periodically to discover and maintain OSPFv3 neighbor relationships.
Database Description (DD) packet	DD packets contain the summary of the local LSDB and are exchanged between two OSPFv3 routers to update the LSDBs.

 Copyright © Huawei Technologies Co., Ltd.

Packet Type	Function
Link State Request (LSR) packet	LSR packets are sent to the neighbor to request the required LSAs. An OSPFv3 device sends LSR packets to its neighbor only after they exchange DD packets.
Link State Update (LSU) packet	LSU packets carry the LSAs required by neighbors.
Link State Acknowledgment (LSAck) packet	LSAck packets acknowledge the receipt of an LSA.

LSA Types

LSA Type	Description
Router-LSA (Type 1)	Describes the link status and link cost of a router. It is generated by every router and advertised in the area to which the router belongs.
Network-LSA (Type 2)	Describes the link status of all routers on the local network segment. Network-LSAs are generated by a designated router (DR) and advertised in the area to which the DR belongs.
Inter-Area-Prefix-LSA (Type 3)	Describes routes to a specific network segment in an area. Inter-Area-Prefix-LSAs are generated on the Area Border Router (ABR) and sent to related areas.
Inter-Area-Router-LSA (Type 4)	Describes routes to an Autonomous System Boundary Router (ASBR). Inter-Area-Router-LSAs are generated by an ABR and advertised to all related areas except the area to which the ASBR belongs.
AS-external-LSA (Type 5)	Describes routes to a destination outside the AS. AS-external-LSAs are generated by an ASBR and advertised to all areas except stub areas and NSSAs.
Link-LSA (Type 8)	Describes the link-local address and IPv6 address prefix associated with the link and the link option set in the network LSA. Link LSAs are transmitted only on the link.

 Copyright © Huawei Technologies Co., Ltd.

LSA Type	Description
Intra-Area-Prefix-LSA (Type 9)	Each device or DR generates one or more intra-area prefix LSAs and transmits it in the local area. • An intra-area prefix LSA generated by a device describes the IPv6 address prefix associated with the router LSA. • An intra-area prefix LSA generated by a DR describes the IPv6 address prefix associated with the network LSA.

Router Types

Figure 11-130 Router types

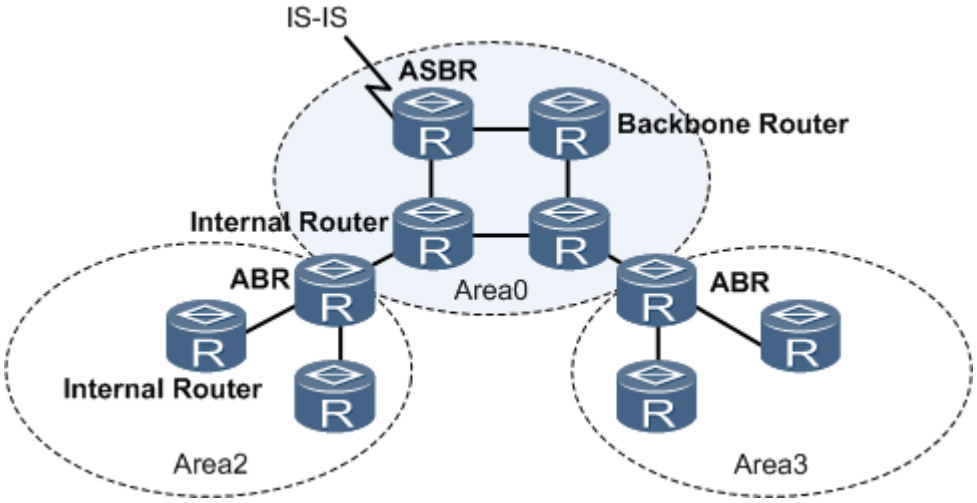


Table 11-57 Router types and descriptions

Router Type	Description
Internal router	All interfaces on an internal router belong to the same OSPFv3 area.
Area border router (ABR)	An ABR belongs to two or more areas, one of which must be the backbone area. An ABR is used to connect the backbone area and non-backbone areas. It can be physically or logically connected to the backbone area.
Backbone router	At least one interface on a backbone router belongs to the backbone area. Internal routers in Area 0 and all ABRs are backbone routers.

 Copyright © Huawei Technologies Co., Ltd.

Router Type	Description
AS boundary router (ASBR)	An ASBR exchanges routing information with other ASs. An ASBR does not necessarily reside on the border of an AS. It can be an internal router or an ABR. An OSPFv3 device that has imported external routing information will become an ASBR.

OSPFv3 Route Types

Inter-area routes and intra-area routes describe the network structure of an AS. External routes describe how to select a route to the destination outside an AS. OSPFv3 classifies the imported AS external routes into Type 1 routes and Type 2 routes.

Table 11-58 lists route types in descending order of priority.

Table 11-58 Types of OSPFv3 routes

Route Type	Description
Intra area route	Indicates routes within an area.
Inter area route	Indicates routes between areas.
Type 1 external route	Type 1 external routes have high reliability. Cost of a Type 1 external route = Cost of the route from a local router to an ASBR + Cost of the route from the ASBR to the destination of the Type 1 external route
Type 2 external route	Type 2 external routes have low reliability, and therefore OSPFv3 considers the cost of the route from an ASBR to the destination of a Type 2 external route to be much greater than the cost of any internal route to the ASBR. Cost of a Type 2 external route = Cost of the route from the ASBR to the destination of the Type 2 external route

Area

When a large number of Routers run OSPFv3, link state databases (LSDBs) become very large and require a large amount of storage space. Large LSDBs also complicate shortest path first (SPF) computation and are computationally intensive for the Routers. Network expansion causes the network topology to change, which results in route flapping and frequent OSPFv3 packet transmission. When a large number of OSPFv3 packets are transmitted on the network,

 Copyright © Huawei Technologies Co., Ltd.

bandwidth usage efficiency decreases. Each change in the network topology causes all Routers on the network to recalculate routes.

OSPFv3 resolves this problem by partitioning an AS into different areas. An area is regarded as a logical group, and each group is identified by an area ID. A Router, not a link, resides at the border of an area. A network segment or link can belong only to one area. An area must be specified for each OSPFv3 interface.

OSPFv3 areas include common areas, stub areas, and not-so-stubby areas (NSSAs), as described in **Table 11-59**.

Table 11-59 OSPFv3 areas

Area Type	Function	Notes
Common area	By default, OSPFv3 areas are defined as common areas. Common areas include: • Standard area: transmits intra-area, inter-area, and external routes. • Backbone area: connects to all other OSPFv3 areas and transmits inter-area routes. The backbone area is represented by area 0. Routes between non-backbone areas must be forwarded through the backbone area.	• The backbone area must have all its devices connected. • All non-backbone areas must remain connected to the backbone area.
Stub area	A stub area is a non-backbone area with only one ABR and generally resides at the border of an AS. The area border router (ABR) in a stub area does not transmit received AS external routes, which significantly decreases the number of entries in the routing table on the ABR and the amount of routing information to be transmitted. To ensure the reachability of AS external routes, the ABR in the stub area generates a default route and advertises the route to non-ABRs in the stub area. A totally stub area allows only intra-area routes and ABR-advertised Type 3 link state advertisements (LSAs) carrying a default route to be advertised within the area.	• The backbone area cannot be configured as a stub area. • An autonomous system boundary router (ASBR) cannot exist in a stub area. Therefore, AS external routes cannot be advertised within the stub area. • A virtual link cannot pass through a stub area.

 Copyright © Huawei Technologies Co., Ltd.

Area Type	Function	Notes
NSSA	An NSSA is similar to a stub area. An NSSA does not advertise Type 5 LSAs, but can import AS external routes. ASBRs in an NSSA generate Type 7 LSAs to carry the information about the AS external routes. The Type 7 LSAs are advertised only within the NSSA. When the Type 7 LSAs reach an ABR in the NSSA, the ABR translates the Type 7 LSAs into Type 5 LSAs and floods them to the entire AS. A totally NSSA area allows only intra-area routes to be advertised within the area.	• ABRs in an NSSA advertise Type 3 LSAs carrying a default route within the NSSA. All inter-area routes are advertised by ABRs. • A virtual link cannot pass through an NSSA.

Network Types Supported by OSPFv3

OSPFv3 classifies networks into the following types based on link layer protocols.

Table 11-60 Types of OSPFv3 networks

Network Type	Description
Broadcast	OSPFv3 considers networks with Ethernet or Fiber Distributed Data Interface (FDDI) as the link layer protocol as broadcast networks by default. On a broadcast network: • Hello packets, LSU packets, and LSAck packets are usually transmitted in multicast mode. FF02::5 is an IPv6 multicast address reserved for an OSPFv3 device. FF02::6 is an IPv6 multicast address reserved for an OSPFv3 DR or backup designated router (BDR). • DD and LSR packets are transmitted in unicast mode.
Non-broadcast Multiple Access (NBMA)	OSPFv3 considers networks with X.25 as the link layer protocol as NBMA networks by default. On an NBMA network, protocol packets, such as Hello packets, DD packets, LSR packets, LSU packets, and LSAck packets are sent in unicast mode.

 Copyright © Huawei Technologies Co., Ltd.

Network Type	Description
Point-to-Multipoint (P2MP)	No network is a P2MP network by default, no matter what type of link layer protocol is used on the network. A non-fully meshed NBMA network can be changed to a P2MP network. On a P2MP network: • Hello packets are transmitted in multicast mode using the multicast address FF02::5. • Other types of protocol packets, such as DD packets, LSR packets, LSU packets, and LSAck packets are sent in unicast mode.
Point-to-point (P2P)	OSPFv3 considers networks with PPP, HDLC, or LAPB as the link layer protocol to be P2P networks by default. On a P2P network, protocol packets, such as Hello packets, DD packets, LSR packets, LSU packets, and LSAck packets are sent in multicast mode using the multicast address FF02::5.

Stub Area

Stub areas are specific areas where ABRs do not flood received AS external routes. In stub areas, routers maintain fewer routing entries and less routing information than those in other areas.

Configuring a stub area is optional. Not every area can be configured as a stub area, because a stub area is usually a non-backbone area with only one ABR and is located at the AS border.

To ensure the reachability of the routes to destinations outside an AS, the ABR in the stub area generates a default route and advertises the route to the non-ABRs in the same stub area.

Note the following points when configuring a stub area:

- The backbone area cannot be configured as a stub area.
- Configure stub area attributes on all routers in the area to be configured as a stub area.
- No ASBRs are allowed in the area to be configured as a stub area because AS external routes cannot be transmitted in the stub area.

OSPFv3 Multi-process

OSPFv3 supports multi-process. Multiple OSPFv3 processes can independently run on the same router. Route exchange between different OSPFv3 processes is similar to that between different routing protocols.

11.12.7.2.2 Comparison Between OSPFv3 and OSPFv2

OSPFv3 and OSPFv2 are the same in the following aspects:

 Copyright © Huawei Technologies Co., Ltd.

- Network types and interface types
- Interface state machines and neighbor state machines
- LSDB
- Flooding mechanism
- Five types of packets: Hello, DD, LSR, LSU, and LSAck packets
- Route calculation

OSPFv3 and OSPFv2 are different in the following aspects:

- In OSPFv3, only LSUs contain IP addresses.
- OSPFv3 runs on IPv6 which is based on links rather than network segments.

  Therefore, the interfaces on which OSPFv3 is to be configured should be on the same link rather than in the same network segment. In addition, the interfaces can establish OSPFv3 sessions without IPv6 global addresses.
- OSPFv3 does not depend on IP addresses.

  OSPFv3 separates topology calculation from IP addresses. Specifically, OSPFv3 can calculate the OSPFv3 topology without IPv6 global addresses which only apply to virtual link interfaces and packet forwarding.
- OSPFv3 packets and the LSA format change.
  - OSPFv3 packets do not contain IP addresses.
  - OSPFv3 router LSAs and network LSAs do not contain IP addresses, which are advertised through link LSAs and intra-area prefix LSAs.
  - In OSPFv3, Router IDs, area IDs, and LSA link state IDs no longer indicate IP addresses, but the IPv4 address format is still reserved.
  - Neighbors are identified by Router IDs instead of IP addresses on broadcast, NBMA, or P2MP networks.
- Information about the flooding scope is added to OSPFv3 LSAs.

  Information about the flooding scope is added to the LSA Type field of OSPFv3 LSAs. Therefore, OSPFv3 routers can process LSAs of unidentified types more flexibly.
  - OSPFv3 can store or flood unidentified packets, while OSPFv2 discards unidentified packets.
  - In OSPFv3, unknown LSAs with 1 as the U flag bit can be flooded, and the flooding scope of such LSAs is specified by the LSAs.

  For example, Router A and Router B can identify LSAs of a certain type. Router A and Router B are connected through Router C which, however, cannot identify these LSAs. If Router A floods such LSA to Router C, Router C can still flood the received LSAs to Router B although Router C does not identify these LSAs. Router B then processes these LSAs.

  If OSPFv2 is run, Router C discards the unidentified LSAs. As a result, these LSAs cannot reach Router B.
- OSPFv3 supports multi-process on a link.

  In OSPFv2, one physical interface can be bound to only one multi-instance. In OSPFv3, one physical interface can be bound to multiple multi-instances that are identified by different instance IDs. In these OSPFv3 multi-instances running on one physical interface, neighbor relationships are established separately, sharing resources on the same link.

 Copyright © Huawei Technologies Co., Ltd.

- OSPFv3 uses IPv6 link-local addresses.

  IPv6 implements neighbor discovery and automatic configuration based on link-local addresses. Routers running IPv6 do not forward IPv6 packets whose destination address is a link-local address, and those packets can only be exchanged on the same link. The unicast link-local address starts from FE80/10.

  As a routing protocol running on IPv6, OSPFv3 also uses link-local addresses to maintain neighbor relationships and update LSDBs. Except Vlink interfaces, all OSPFv3 interfaces use link-local addresses as the source address and the next hop to transmit OSPFv3 packets.

  The advantages are as follows:

  – OSPFv3 can calculate the topology without global IPv6 addresses.

  – The packets flooded on a link are not transmitted to other links, which prevents unnecessary flooding and saves bandwidth.

- OSPFv3 supports two new LSAs.

  – Link LSA: A device floods a link LSA on the link where it resides to advertise its link-local address and the configured global IPv6 address.

  – Intra-area prefix LSA: A device advertises an intra-area prefix LSA in the local OSPF area to inform the other routers in the area or the network (either a broadcast network or an NBMA network) of its IPv6 global address.

- OSPFv3 identifies neighbors based on Router IDs only.

  On broadcast, NBMA, and P2MP networks, OSPFv2 identifies neighbors based on IPv4 addresses of interfaces.

  OSPFv3 identifies neighbors based on Router IDs only.

11.12.7.2.3 OSPFv3 GR

NOTE

The can be configured as a GR helper rather than a GR restarter.

Graceful restart (GR) is a technology used to ensure proper traffic forwarding, especially the forwarding of key services, during the restart of routing protocols.

Without GR, the master/slave main control board switchover due to various reasons leads to transient service interruption, and as a result, route flapping occurs on the whole network. Such route flapping and service interruption are unacceptable on large-scale networks, especially carrier networks.

GR is one of the high availability (HA) technologies which comprise a series of comprehensive technologies, such as fault-tolerant redundancy, link protection, faulty node recovery, and traffic engineering technologies. As a fault-tolerant redundancy technology, GR is widely used to ensure non-stop forwarding of key data during the master/slave main control board switchovers and system upgrade.

In GR mode, the forwarding plane continues data forwarding during a restart, and operations on the control plane, such as re-establishment of neighbor relationships and route calculation, do not affect the forwarding plane, preventing service interruptions caused by route flapping and improving network reliability.

 Copyright © Huawei Technologies Co., Ltd.

Comparison Between Master/Slave Main Control Board Switchovers with and Without GR

Table 11-61 Comparison between master/slave main control board switchovers with and without GR

Master/Slave Main Control Board Switchovers Without GR	Master/Slave Main Control Board Switchovers with GR
• OSPFv3 neighbor relationships are reestablished. • Routes are recalculated. • The forwarding table changes. • The entire network detects route changes, and route flapping occurs for a short period of time. • Packets are lost during forwarding, and services are interrupted.	• OSPFv3 neighbor relationships are reestablished. • Routes are recalculated. • The forwarding table remains unchanged. • Except the neighbors of the router on which a master/slave main control board switchover occurs, other routers do not detect route changes. • No packets are lost during forwarding, and services are not affected.

11.12.7.2.4 OSPFv3 Authentication

In Open Shortest Path First version 3 (OSPFv3) authentication, an authentication field is added to each OSPFv3 packet for encryption. When a local device receives an OSPFv3 packet from a remote device, the local device discards the packet if the authentication password carried in the packet is different from the local one, which protects the local device against potential attacks. Therefore, OSPFv3 authentication improves network security.

Based on the applicable scope, OSPFv3 authentication is classified as follows:

- Area authentication

  This authentication is configured in the OSPFv3 area view and applies to packets received by all interfaces in an OSPF area.
- Process authentication

  This authentication is configured in the OSPFv3 view and applies to all packets in an OSPF process.
- Interface authentication

  This authentication is configured in the interface view and applies to all packets received by the interface.

OSPFv3 uses HMAC-SHA256 authentication to authenticate packets. In HMAC-SHA256 authentication, a password is encrypted using the HMAC-SHA256 algorithm before being added to a packet, which improves password security.

Each OSPFv3 packet carries an authentication type in the header and authentication information in the tail.

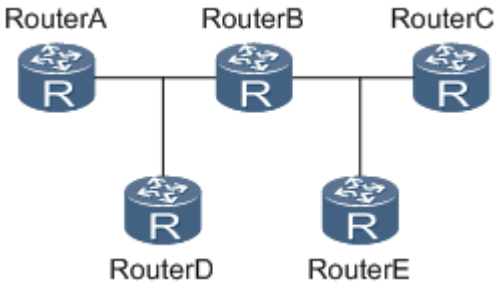
The authentication types are as follows:

 Copyright © Huawei Technologies Co., Ltd.

- 1: simple authentication
- 2: ciphertext authentication

Networking Application

Figure 11-131 OSPFv3 authentication on a broadcast network



The configuration requirements are as follows:

- Interface authentication configurations must be the same on all devices of the same network so that OSPFv3 neighbor relationships can be established.
- Area authentication configurations must be the same on all devices in the same area.

11.12.7.2.5 OSPFv3 Packet Format

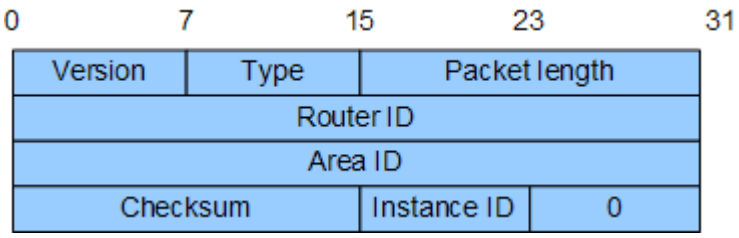
Open Shortest Path First (OSPF) for IPv6 packets are encapsulated into IPv6 packets. The OSPFv3 protocol number is 89. OSPFv3 packets are classified into the following types:

- **Hello packet**
- **Database Description (DD) packet**
- **Link State Request (LSR) packet**
- **Link State Update (LSU) packet**
- **Link State Acknowledgment (LSAck) packet**

Packet Header Format

The five types of OSPFv3 packets have the same packet header format. The length of an OSPFv3 packet header is 24 bytes. **Figure 11-132** shows an OSPFv3 packet header.

Figure 11-132 OSPFv3 packet header



 Copyright © Huawei Technologies Co., Ltd.

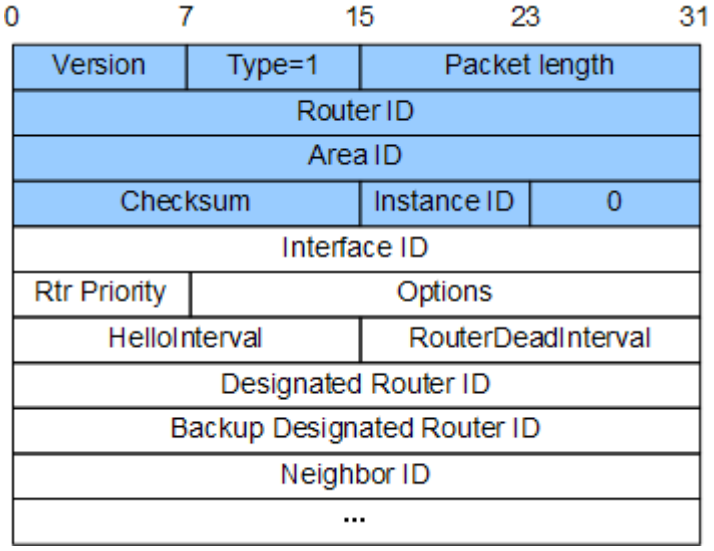
Table 11-62 Packet header fields

Field	Length	Description
Version	8 bits	OSPF version number. For OSPFv3, the value is 3.
Type	8 bits	OSPFv3 packet type. The values are as follows: • 1: Hello packet • 2: DD packet • 3: LSR packet • 4: LSU packet • 5: LSAck packet
Packet length	16 bits	Length of the OSPFv3 packet containing the packet header, in bytes.
Router ID	32 bits	ID of the Router that sends the OSPFv3 packet.
Area ID	32 bits	ID of the area to which the Router that sends the OSPFv3 packet belongs.
Checksum	16 bits	Checksum of the OSPFv3 packet that does not contain the Authentication field.
Instance ID	8 bits	ID of an OSPFv3 instance.
0	8 bits	Reserved fields.

Hello Packet

Hello packets are commonly used packets, which are periodically sent on OSPFv3 interfaces to establish and maintain neighbor relationships. A Hello packet includes information about the designated router (DR), backup designated router (BDR), timers, and known neighbors. **Figure 11-133** shows the format of a Hello packet.

Figure 11-133 Format of a Hello packet



 Copyright © Huawei Technologies Co., Ltd.

Table 11-63 Hello packet fields

Field	Length	Description
Interface ID	32 bits	ID of the interface that sends the Hello packets.
Rtr Priority	8 bits	DR priority. The default value is 1. **NOTE** If the DR priority of a Router interface is set to 0, the interface cannot participate in a DR or BDR election.
Options	24 bits	The values are as follows: • E: Type 5 link state advertisements (LSAs) are flooded. • MC: IP multicast packets are forwarded. • N/P: Type 7 LSAs are processed. • DC: On-demand links are processed.
HelloInterval	16 bits	Interval at which Hello packets are sent.
RouterDeadInterval	16 bits	Dead interval. If a Router does not receive any Hello packets from its neighbors within a specified dead interval, the neighbors are considered Down.
Designated Router ID	32 bits	Interface address of the DR.
Backup Designated Router ID	32 bits	Interface address of the BDR.
Neighbor ID	32 bits	Router ID of the neighbor.

Table 11-64 lists the address types, interval types, and default intervals used when Hello packets are transmitted on different networks.

Table 11-64 Hello packet characteristics for various network types

Network Type	Address Type	Interval Type	Default Interval
Broadcast	Multicast address	HelloInterval	10 seconds

 Copyright © Huawei Technologies Co., Ltd.

Network Type	Address Type	Interval Type	Default Interval
Non-broadcast multiple access (NBMA)	Unicast address	• HelloInterval for the DR, BDR, and Router that can become a DR • PollInterval for the case when neighbors become Down and HelloInterval for other cases	30 seconds for HelloInterval 120 seconds for PollInterval
Point-to-point (P2P)	Multicast address	HelloInterval	10 seconds
Point-to-multipoint (P2MP)	Unicast address	HelloInterval	30 seconds

NOTE

To establish neighbor relationships between Routers on the same network segment, you must set the same HelloInterval, PollInterval, and RouterDeadInterval values for the Routers. PollInterval applies only to NBMA networks.

DD Packet

During an adjacency initialization, two Routers use DD packets to describe their own link state databases (LSDBs) for LSDB synchronization. A DD packet contains the header of each LSA in an LSDB. An LSA header uniquely identifies an LSA. The LSA header occupies only a small portion of the LSA, which reduces the amount of traffic transmitted between Routers. A neighbor can use the LSA header to check whether it already has the LSA. When two Routers exchange DD packets, one functions as the master and the other functions as the slave. The master defines a start sequence number. The master increases the sequence number by one each time it sends a DD packet. After the slave receives a DD packet, it uses the sequence number carried in the DD packet for acknowledgement.

Figure 11-134 shows the format of a DD packet.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-134 Format of a DD packet

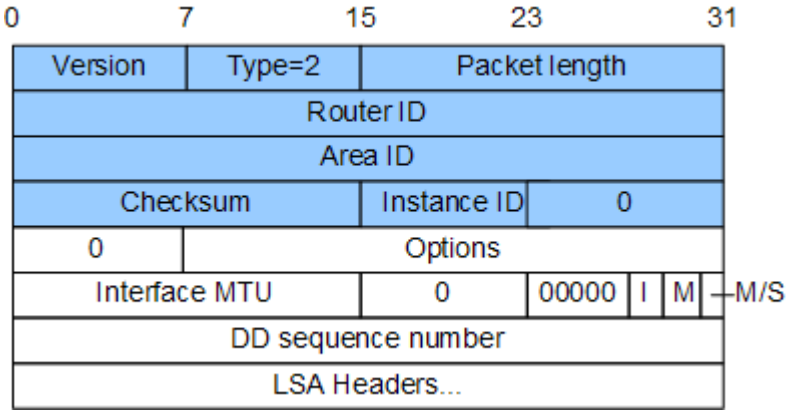


Table 11-65 DD packet fields

Field	Length	Description
Options	24 bits	The values are as follows: • E: Type 5 LSAs are flooded. • MC: IP multicast packets are forwarded. • N/P: Type 7 LSAs are processed. • DC: On-demand links are processed.
Interface MTU	16 bits	Maximum length of the DD packet sent by the interface with packet fragmentation disabled.
I	1 bit	If the DD packet is the first packet among multiple consecutive DD packets sent by a Router, this field is set to 1. In other cases, this field is set to 0.
M (More)	1 bit	If the DD packet is the last packet among multiple consecutive DD packets sent by a Router, this field is set to 0. In other cases, this field is set to 1.
M/S (Master/ Slave)	1 bit	When two Routers exchange DD packets, they negotiate a master/slave relationship. The Router with a larger router ID becomes the master. If this field is set to 1, the DD packet is sent by the master.
DD sequence number	32 bits	Sequence number of the DD packet. The master and slave use the sequence number to ensure that DD packets are correctly transmitted.
LSA Headers	-	LSA header information included in the DD packet.

LSR Packet

After two Routers exchange DD packets, they send LSR packets to request each other's LSAs. The LSR packets contain the summaries of the requested LSAs. **Figure 11-135** shows the format of an LSR packet.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-135 Format of an LSR packet

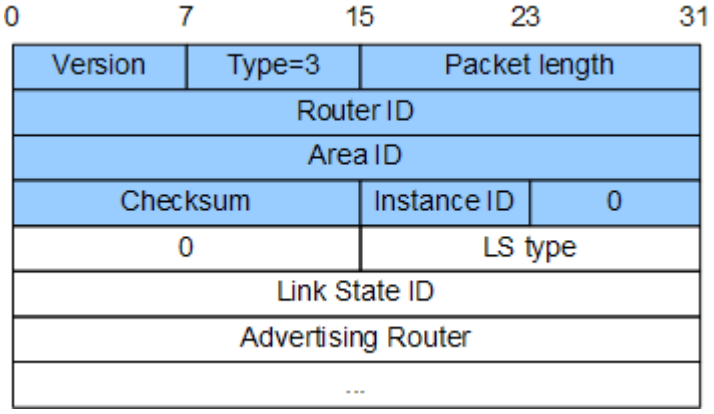


Table 11-66 LSR packet fields

Field	Length	Description
LS type	16 bits	Type of the LSA
Link State ID	32 bits	This field together with the LS type field describes an LSA in an AS.
Advertising Router	32 bits	Router ID of the Router that generates the LSA.

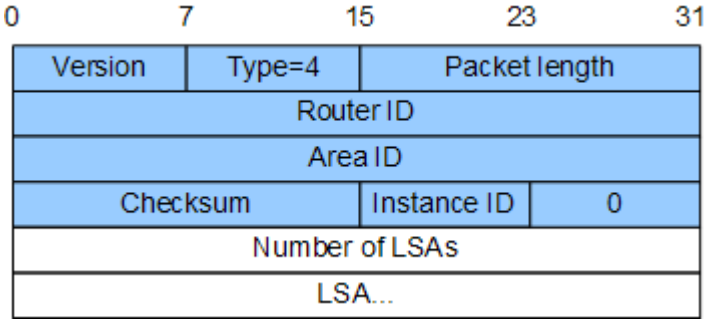
NOTE

The LS type, Link State ID, and Advertising Router fields can uniquely identify an LSA. If two LSAs have the same LS type, Link State ID, and Advertising Router fields, a Router uses the LS sequence number, LS checksum, and LS age fields to obtain a required LSA.

LSU Packet

A Router uses an LSU packet to transmit LSAs requested by its neighbors or to flood its own updated LSAs. The LSU packet contains a set of LSAs. For multicast and broadcast networks, LSU packets are multicast to flood LSAs. To ensure reliable LSA flooding, a Router uses an LSAck packet to acknowledge the LSAs contained in an LSU packet that is received from a neighbor. If an LSA fails to be acknowledged, the Router retransmits the LSA to the neighbor. **Figure 11-136** shows the format of an LSU packet.

Figure 11-136 Format of an LSU packet



 Copyright © Huawei Technologies Co., Ltd.

Table 11-67 LSU packet field

Field	Length	Description
Number of LSAs	32 bits	Number of LSAs contained in the LSU packet

LSAck Packet

A Router uses an LSAck packet to acknowledge the LSAs contained in a received LSU packet. The LSAs can be acknowledged using LSA headers. LSAck packets can be transmitted over different links in unicast or multicast mode. **Figure 11-137** shows the format of an LSAck packet.

Figure 11-137 Format of an LSAck packet

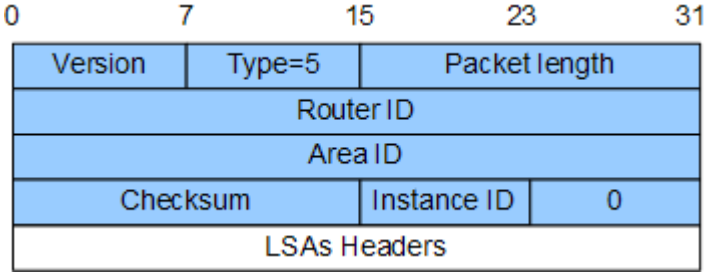


Table 11-68 LSAck packet field

Field	Length	Description
LSAs Headers	Determined by the header length of the LSA to be acknowledged.	This field is used to acknowledge an LSA.

11.12.7.2.6 OSPFv3 LSA Format

Each Router in an autonomous system (AS) generates one or more types of link state advertisements (LSAs), depending on the Router's type. Multiple LSAs form a link state database (LSDB). Open Shortest Path First (OSPF) for IPv6 encapsulates routing information into LSAs for transmission. Commonly used LSAs include:

- **Router-LSA (Type1)**
- **Network-LSA (Type2)**
- **Inter-Area-Prefix-LSA (Type3)**
- **Inter-Area-Router-LSA (Type4)**
- **AS-external-LSA (Type5)**
- **NSSA LSA (Type7)**

 Copyright © Huawei Technologies Co., Ltd.

- **Link-LSA (Type8)**
- **Intra-Area-Prefix-LSA (Type9)**

LSA Header Format

All LSAs have the same header. **Figure 11-138** shows an LSA header.

Figure 11-138 LSA header

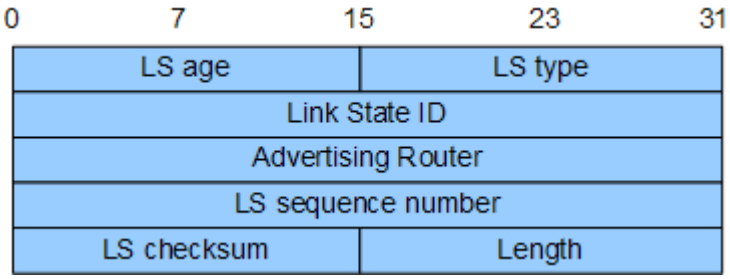


Table 11-69 LSA header fields

Field	Length	Description
LS age	16 bits	Time that elapses after the LSA is generated, in seconds. The value of this field continually increases regardless of whether the LSA is transmitted over a link or saved in an LSDB.
LS type	16 bits	Type of the LSA. The values are as follows: • Type1: Router-LSA. • Type2: Network-LSA. • Type3: Inter-Area-Prefix-LSA. • Type4: Inter-Area-Router-LSA. • Type5: AS-external-LSA. • Type7: NSSA-LSA. • Type8: Link-LSA. • Type9: Intra-Area-Prefix-LSA.
Link State ID	32 bits	This field together with the LS type field describes an LSA in an AS.
Advertising Router	32 bits	Router ID of the Router that generates the LSA.
LS sequence number	32 bits	Sequence number of the LSA. Neighbors can use this field to identify the latest LSA.
LS checksum	16 bits	Checksum of all fields except the LS age field.
Length	16 bits	Length of the LSA including the LSA header, in bytes.

 Copyright © Huawei Technologies Co., Ltd.

Router-LSA

A router-LSA describes the link status and cost of a Router. Router-LSAs are generated by a Router and advertised within the area to which the Router belongs. **Figure 11-139** shows the format of a router-LSA.

Figure 11-139 Format of a router-LSA

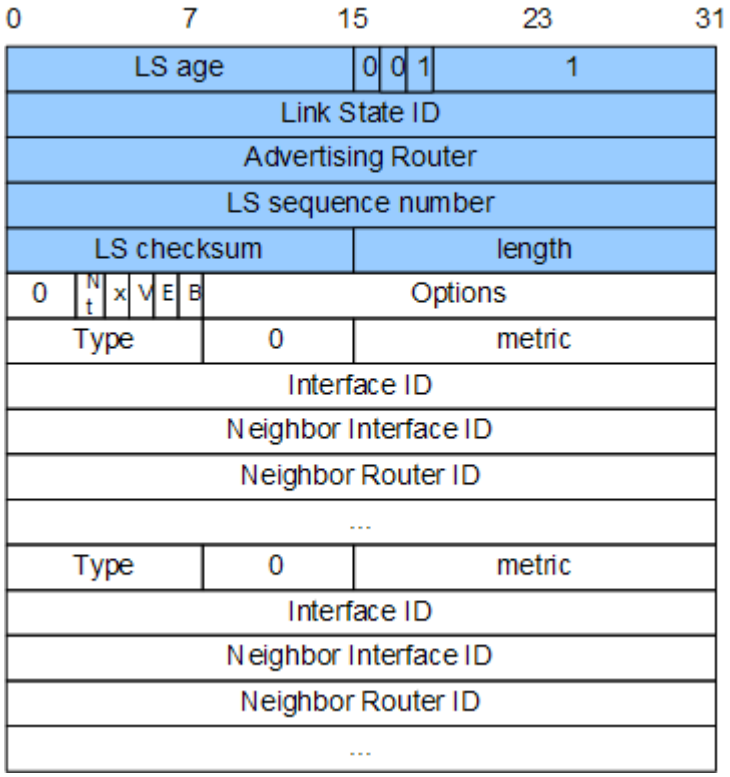


Table 11-70 Router-LSA fields

Field	Length	Description
Nt (NSSA translation)	1 bit	If the Router that generates the LSA is an NSSA border router, this field is set to 1. In other cases, this field is set to 0. When this field is set to 1, the router is unconditionally translating NSSA-LSAs into AS-external-LSAs.
x	1 bit	This field is deprecated.
V (Virtual Link)	1 bit	If the Router that generates the LSA is located at one end of a virtual link, this field is set to 1. In other cases, this field is set to 0.
E (External)	1 bit	If the Router that generates the LSA is an autonomous system boundary router (ASBR), this field is set to 1. In other cases, this field is set to 0.

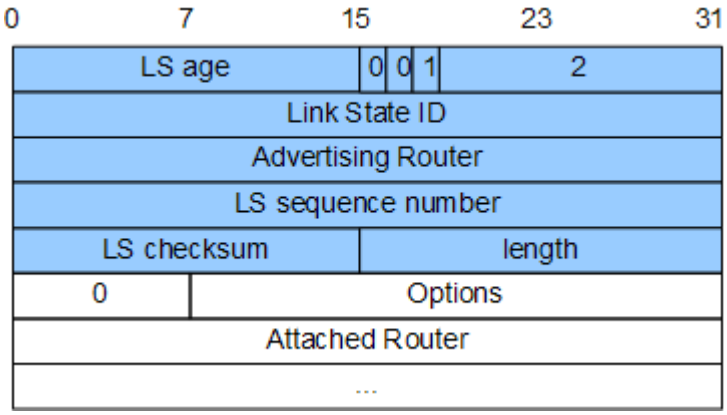
Copyright © Huawei Technologies Co., Ltd.

Field	Length	Description
B (Border)	1 bit	If the Router that generates the LSA is an area border router (ABR), this field is set to 1. In other cases, this field is set to 0.
Options	24 bits	The optional capabilities supported by the Router.
Type	8 bits	Type of the Router link. The values are as follows: • 1: The Router is connected to another Router in point-to-point (P2P) mode. • 2: The Router is connected to a transport network. • 3: Reserved. • 4: The Router is connected to another Router over a virtual link.
metric	16 bits	Cost of the link.
Interface ID	32 bits	The Interface ID assigned to the interface.
Neighbor Interface ID	32 bits	The Interface ID the neighbor Router has associated with the link.
Neighbor Router ID	32 bits	The Router ID of the neighbor Router.

Network-LSA

A network-LSA describes the link status of all Routers on the local network segment. Network-LSAs are generated by a DR on a broadcast or non-broadcast multiple access (NBMA) network and advertised within the area to which the DR belongs. **Figure 11-140** shows the format of a network-LSA.

Figure 11-140 Format of a network-LSA



 Copyright © Huawei Technologies Co., Ltd.

Table 11-71 Network-LSA fields

Field	Length	Description
Options	24 bits	The optional capabilities supported by the Router.
Attached Router	32 bits	Router IDs of all Routers on the broadcast or NBMA network, including the router ID of the DR

Inter-Area-Prefix-LSA

An inter-area-prefix-LSA describes routes on a network segment in an area. It is generated by the ABR. The routes are advertised to other areas.

Figure 11-141 shows the format of an inter-area-prefix-LSA.

Figure 11-141 Format of An inter-area-prefix-LSA

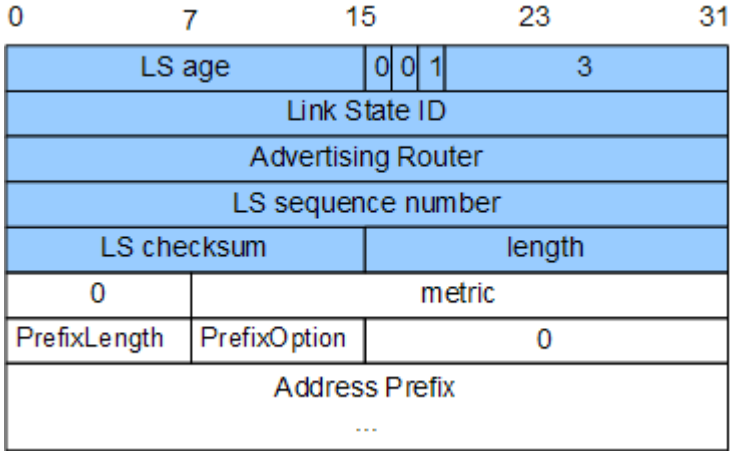


Table 11-72 Network-summary-LSA fields

Field	Length	Description
PrefixLength	8 bits	Length of the prefix.
PrefixOption	8 bits	The capabilities associated with the prefix.
Address Prefix	32 bits	An encoding of the prefix itself.

Inter-Area-Router-LSA

An inter-area-router-LSA describes routes to ASBR in other areas. It is generated by the ABR. The routes are advertised to all related areas except the area that the ASBR belongs to.

Figure 11-142 shows the format of an inter-area-prefix-LSA.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-142 Format of an inter-area-router-LSA

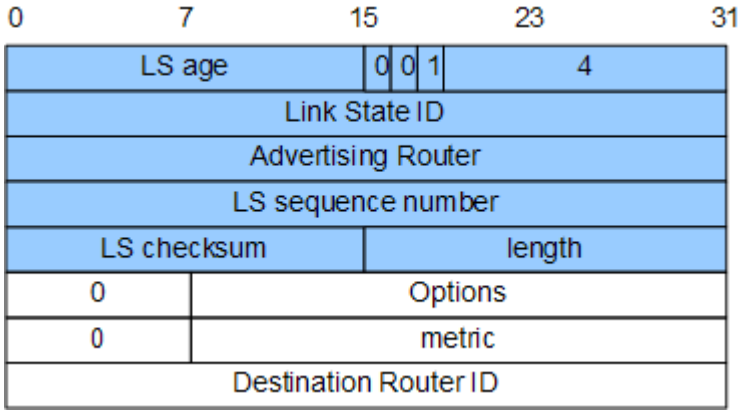


Table 11-73 Inter-area-router-LSA fields

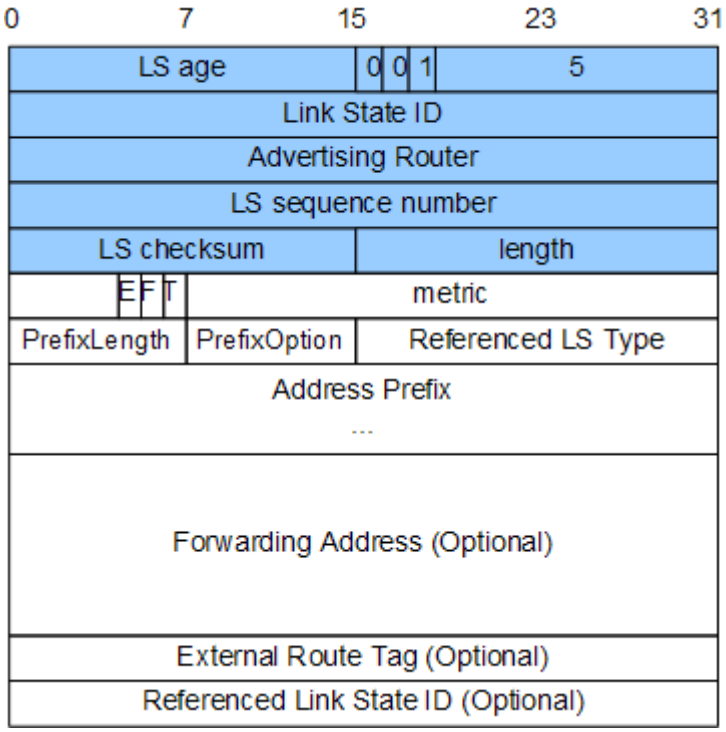
Field	Length	Description
Destination Router ID	32 bits	The Router ID of the router being described by the LSA.

AS-External-LSAs

An as-external-LSA describes destinations outside the AS, it is originated by ASBR.

Figure 11-143 shows the format of an as-external-LSA.

Figure 11-143 Format of an as-external-LSA



 Copyright © Huawei Technologies Co., Ltd.

Table 11-74 As-external-LSA fields

Field	Length	Description
E	1 bit	The type of external metric. • If this field is 1, the specified metric is a Type 2 external metric. • If this field is 0, the specified metric is a Type 1 external metric.
F	1 bit	Whether a Forwarding Address has been included in the LSA. • If this field is 1, a Forwarding Address has been included in the LSA. • If this field is 0, no Forwarding Address is included in the LSA.
T	1 bit	Whether an External Route Tag has been included in the LSA. • If this field is 1, an External Route Tag has been included in the LSA. • If this field is 0, no External Route Tag is included in the LSA.
Referenced LS Type	16 bits	Indicates the referenced LS type. If non-zero, an LSA with this LS type is to be associated with this LSA (see Referenced Link State ID below).
Forwarding Address	128 bits	A fully qualified global IPv6 address.
External Route Tag	32 bits	Indicates the external route tag, which can be used to communicate additional information between ASBRs.
Referenced Link State ID	32 bits	Indicates the referenced link state ID.

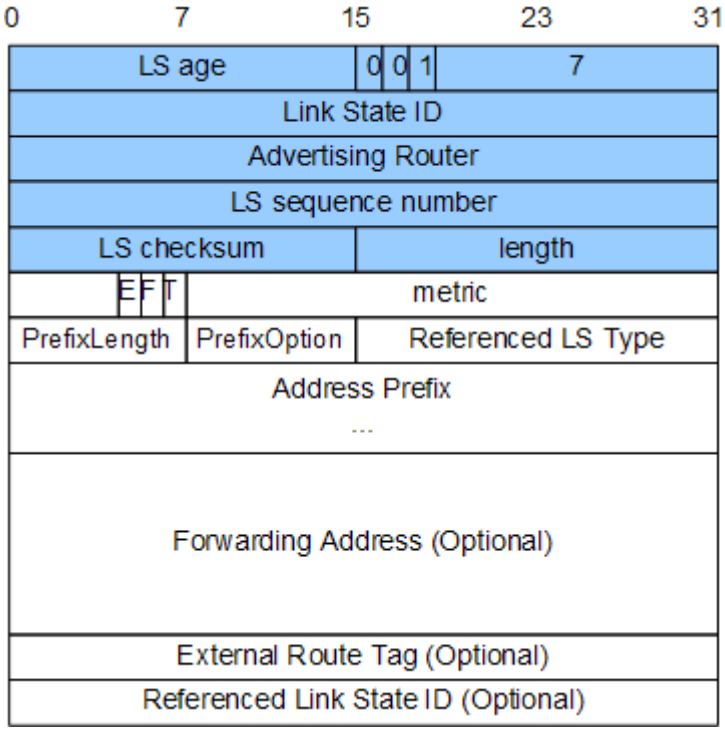
NSSA-LSAs

An NSSA-LSA describes destinations outside the AS, it is originated by ASBR.

Figure 11-144 shows the format of an NSSA-LSA.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-144 Format of an NSSA-LSA



Link-LSAs

Each Router generates a link LSA for each link. A link LSA describes the link-local address and IPv6 address prefix associated with the link and the link option set in the network LSA. It is transmitted only on the link.

Figure 11-145 shows the format of a Link-LSA.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-145 Format of a Link-LSA

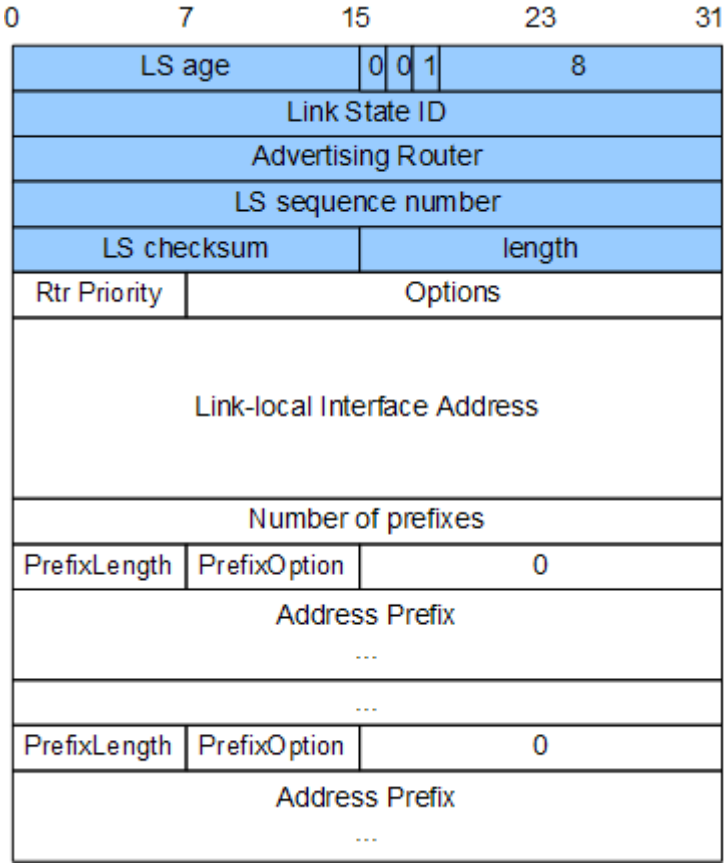


Table 11-75 Link-LSA fields

Field	Length	Description
Rtr Priority	8 bits	The Router Priority of the interface.
Options	24 bits	The set of Options bits that the Router would like set in the network-LSA that will be originated by the DR on broadcast or NBMA links.
Link-local Interface Address	128 bits	The originating Router's link-local interface address on the link.
Number of prefixes	32 bits	The number of IPv6 address prefixes contained in the LSA.

Intra-Area-Prefix-LSAs

Each Router or DR generates one or more intra-area prefix LSAs and transmits it in the local area.

- An LSA generated on a router describes the IPv6 address prefix associated with the router LSA.

 Copyright © Huawei Technologies Co., Ltd.

- An LSA generated on a DR describes the IPv6 address prefix associated with the network LSA.

Figure 11-146 shows the format of an intra-area-prefix-LSA.

Figure 11-146 Format of an intra-area-prefix-LSA

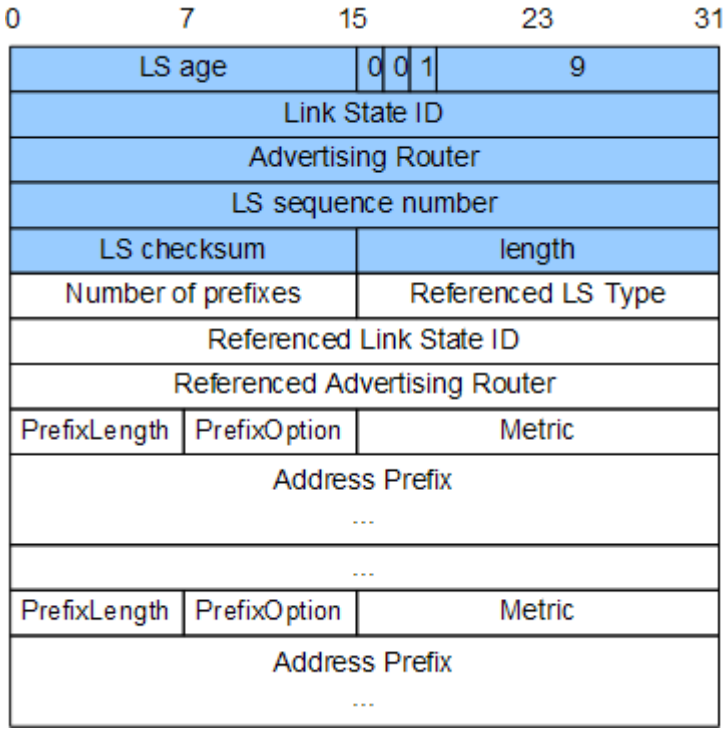


Table 11-76 Intra-area-prefix-LSA fields

Field	Length	Description
Referenced LS Type	16 bits	This field identifies the router-LSA or network-LSA with which the IPv6 address prefixes should be associated. • If Referenced LS Type is 0x2001, the prefixes are associated with a router-LSA. • If Referenced LS Type is 0x2002, the prefixes are associated with a network-LSA.
Referenced Link State ID	32 bits	Indicates the referenced link state ID. • If Referenced LS Type is 0x2001, Referenced Link State ID should be 0. • If Referenced LS Type is 0x2002, Referenced Link State ID should be the Interface ID of the link's DR.

 Copyright © Huawei Technologies Co., Ltd.

Field	Length	Description
Referenced Advertising Router	32 bits	Indicates the ID of the referenced advertising Router. • If Referenced LS Type is 0x2001, Referenced Advertising Router should be the originating router's Router ID. • If Referenced LS Type is 0x2002, Referenced Advertising Router should be the Designated Router's Router ID.

11.12.7.2.7 OSPFv3 Auto FRR

OSPFv3 auto fast reroute (FRR) is a kind of dynamic IP FRR. To implement this feature, Interior Gateway Protocol (IGP) calculates the backup paths beforehand by using the link status database of the entire network and then stores the backup paths in the forwarding table to minimize service interruption duration in case of a link fault.

Background

Some services, such as voice over IP (VoIP) and online video services, are delay-sensitive. Nevertheless, if a primary link fails, OSPFv3 must perform multiple operations, including detecting the fault, updating the link-state advertisement (LSA), flooding the LSA, calculating routes, and delivering forward information base (FIB) entries before switching traffic to a new link. This process takes a much longer time than the delay-sensitive services can tolerate. To solve this problem, OSPFv3 Auto FRR is developed.

Related Concepts

OSPFv3 Auto FRR

OSPFv3 Auto FRR refers to a mechanism in which a device uses the loop-free alternate (LFA) algorithm to compute a backup link and stores it with the primary link in the forwarding table. If the primary link fails, the device switches the traffic to the backup link before routes are converged on the control plane to minimize the impact of the link failure.

NOTE

If a Bidirectional Forwarding Detection (BFD) session is configured, it automatically associates with OSPFv3 auto FRR. If a BFD session detects a link failure, the BFD session goes down and triggers FRR.

OSPFv3 Auto FRR Policy

An OSPFv3 Auto FRR policy can be configured to filter backup routes. Only the backup routes that match the filtering rules of the policy can be added to the IP routing table.

LFA algorithm

A device uses each neighbor that can be used as a backup next hop as a root node, uses the shortest path first (SPF) algorithm to calculate the shortest path

 Copyright © Huawei Technologies Co., Ltd.

from each root node to the destination, and then uses the inequalities defined in RFC 5286 to calculate the backup link that is loop-free and has the lowest cost.

Networking

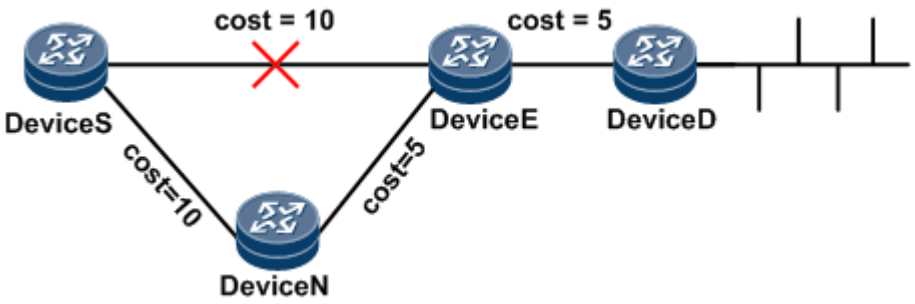
OSPFv3 auto FRR protects traffic from a fault on a specified link.

In **Figure 11-147**, traffic flows from Device S to Device D. The primary link is Device S->Device E->Device D, and the backup link is Device S->Device N->Device E->Device D. The link costs meet the inequality: Distance_opt(N, D) < Distance_opt(N, S) + Distance_opt(S, D). With OSPFv3 Auto FRR, Device S switches the traffic to the backup link if the primary link fails, minimizing the traffic interruption duration.

NOTE

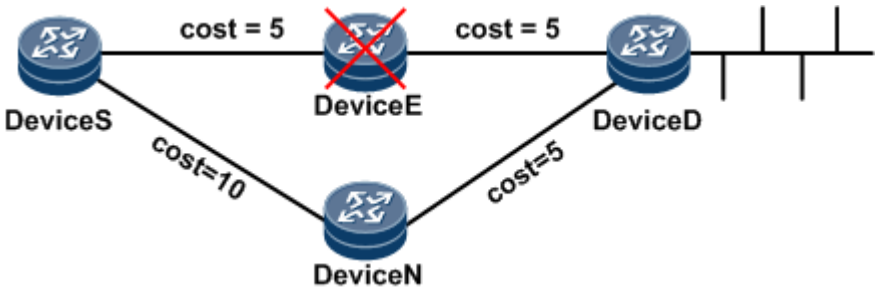
Distance_opt(X, Y) indicates the shortest path from X to Y. S stands for the source node, N for a node along the backup link, and D for the destination node.

Figure 11-147 OSPFv3 Auto FRR link protection



The switching process in the following figure is similar. Therefore, the details are omitted.

Figure 11-148 OSPFv3 Auto FRR node-and-link protection



11.12.7.3 Configuration Example of OSPFv3

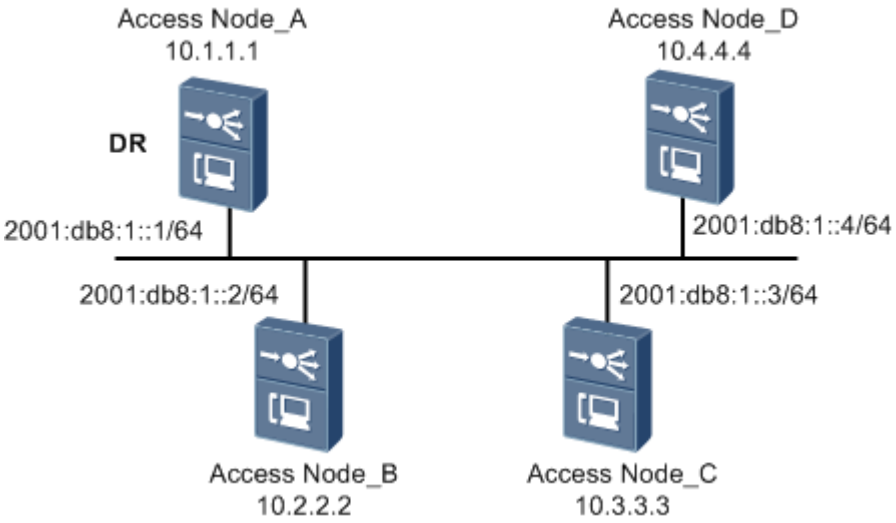
This topic provides an example for configuring OSPFv3 on the MA5800.

 Copyright © Huawei Technologies Co., Ltd.

Service Requirements

- OSPFv3 is enabled on the four MA5800s.
- Access Node_A is configured with the highest designated router (DR) priority, Access Node_C is configured with the second highest DR priority, and Access Node_A implements the broadcast of network link status for the DR.

Figure 11-149 Example network for configuring OSPFv3



Data Plan

Table 11-77 provides the data plan for configuring OSPFv3.

Table 11-77 Data plan for configuring OSPFv3

Item	Data	Remarks
Access Node_A	IPv6 address of the Layer 3 interface: 2001:db8:1::1/64	-
	Priority: 100	Ensure one of the access nodes is configured with the highest designated router (DR) priority. Take Access Node_A for example.
	VLAN ID: 2	-
	Router ID: 10.1.1.1	-
Access Node_B	IPv6 address of the Layer 3 interface: 2001:db8:1::2/64	-
	Priority: 80	-

 Copyright © Huawei Technologies Co., Ltd.

Item	Data	Remarks
	VLAN ID: 2	-
	Router ID: 10.2.2.2	-
Access Node_C	IPv6 address of the Layer 3 interface: 2001:db8:1::3/64	-
	Priority: 90	-
	VLAN ID: 2	-
	Router ID: 10.3.3.3	-
Access Node_D	IPv6 address of the Layer 3 interface: 2001:db8:1::4/64	-
	Priority: not configured	Default: 1
	VLAN ID: 2	-
	Router ID: 10.4.4.4	-

Prerequisite

The OSPFv3 area IDs of the MA5800 devices must be consistent. Take area 0 for example.

Procedure

Step 1 Configure Access Node_A.

1. Configure the IPv6 address of the Layer 3 interface.

```
huawei(config)#vlan 2
huawei(config)#port vlan 2 0/9 0
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ipv6 enable
huawei(config-if-vlanif2)#ipv6 address 2001:db8:1::1 64
huawei(config-if-vlanif2)#quit
```

2. Enable OSPFv3, and configure the OSPFv3 Router ID.

```
huawei(config)#ospfv3
huawei(config-ospfv3-1)#router-id 10.1.1.1
huawei(config-ospfv3-1)#quit
```

3. Enable OSPFv3 on the vlanif interface, and configure the OSPFv3 priority.

```
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ospfv3 1 area 0
huawei(config-if-vlanif2)#ospfv3 dr-priority 100
huawei(config-if-vlanif2)#quit
```

4. Save the data.

```
huawei(config)#save
```

Step 2 Configure Access Node_B.

1. Configure the IPv6 address of the Layer 3 interface.

```
huawei(config)#vlan 2
huawei(config)#port vlan 2 0/9 0
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ipv6 enable
huawei(config-if-vlanif2)#ipv6 address 2001:db8:1::2 64
huawei(config-if-vlanif2)#quit
```

2. Enable OSPFv3, and configure the OSPFv3 Router ID.

```
huawei(config)#ospfv3
huawei(config-ospfv3-1)#router-id 10.2.2.2
huawei(config-ospfv3-1)#quit
```

3. Enable OSPFv3 on the vlanif interface, and configure the OSPFv3 priority.

```
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ospfv3 1 area 0
huawei(config-if-vlanif2)#ospfv3 dr-priority 80
huawei(config-if-vlanif2)#quit
```

4. Save the data.

```
huawei(config)#save
```

Step 3 Configure Access Node_C.

1. Configure the IPv6 address of the Layer 3 interface.

```
huawei(config)#vlan 2
huawei(config)#port vlan 2 0/9 0
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ipv6 enable
huawei(config-if-vlanif2)#ipv6 address 2001:db8:1::3 64
huawei(config-if-vlanif2)#quit
```

2. Enable OSPFv3, and configure the OSPFv3 Router ID.

```
huawei(config)#ospfv3
huawei(config-ospfv3-1)#router-id 10.3.3.3
huawei(config-ospfv3-1)#quit
```

3. Enable OSPFv3 on the vlanif interface, and configure the OSPFv3 priority.

```
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ospfv3 1 area 0
huawei(config-if-vlanif2)#ospfv3 dr-priority 90
huawei(config-if-vlanif2)#quit
```

4. Save the data.

```
huawei(config)#save
```

Step 4 Configure Access Node_D.

1. Configure the IPv6 address of the Layer 3 interface.

```
huawei(config)#vlan 2
huawei(config)#port vlan 2 0/9 0
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ipv6 enable
huawei(config-if-vlanif2)#ipv6 address 2001:db8:1::4 64
huawei(config-if-vlanif2)#quit
```

2. Enable OSPFv3, and configure the OSPFv3 Router ID.

```
huawei(config)#ospfv3
huawei(config-ospfv3-1)#router-id 10.4.4.4
uawei(config-ospfv3-1)#quit
```

3. Enable OSPFv3 on the vlanif interface.

```
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ospfv3 1 area 0
huawei(config-if-vlanif2)#quit
```

4. Save the data.

```
huawei(config)#save
```

----End

 Copyright © Huawei Technologies Co., Ltd.

Result

Run the **display ipv6 routing-table** command and you can find the learnt route table. Hosts can communicate with each other.

Configuration File

Configuration on each MA5800 is similar. Take Access Node_A for example.

```
vlan 2 smart
port vlan 2 0/9 0
interface vlanif 2
ipv6 enable
ipv6 address 2001:db8:1::1 64
quit
ospfv3
router-id 10.1.1.1
quit
interface vlanif 2
ospfv3 1 area 0
ospfv3 dr-priority 100
quit
save
```

11.12.7.4 Example for Configuring OSPFv3 Auto FRR

OSPFv3 auto FRR can generate a loop-free backup link for fast link switching in case of a link failure.

Service Requirements

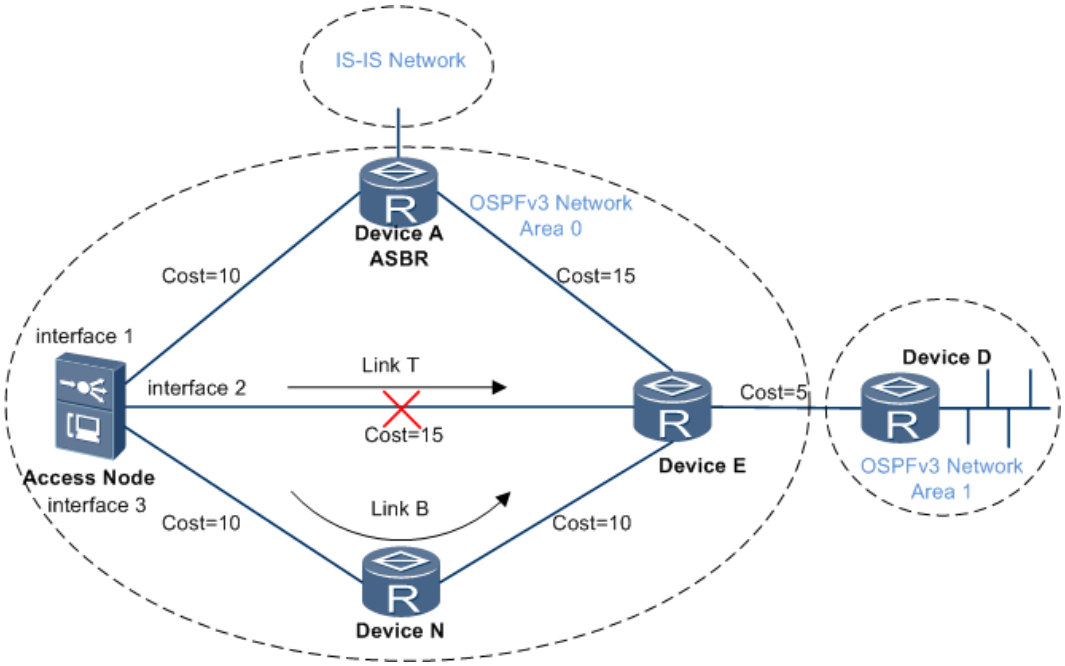
OSPFv3 auto FRR quickly switches traffic to the backup link without waiting for route convergence. Moreover, you can exclude a specified link from backup link options.

As shown in **Figure 11-150**,

- OSPFv3 is running on all devices.
- The link costs meet the OSPFv3 auto FRR traffic protection inequality.
- If the primary link T fails, traffic forwarded by the MA5800 is quickly switched to the backup link B.
- In network planning, you can exclude a link from backup link options.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-150 Networking for configuring OSPFv3 auto FRR



Data Plan

Table 11-78 describes data plan for configuring OSPFv3 auto FRR.

Table 11-78 Data plan for configuring OSPFv3 auto FRR

Item	Data
MA5800 port 1	IPv6 address of the interface: 2001:db8:1000::1 64
	VLAN ID: 2
	Cost: 10
MA5800 port 2	IPv6 address of the interface: 2001:db8:1001::1 64
	VLAN ID: 3
	Cost: 15
MA5800 port 3	IPv6 address of the interface: 2001:db8:1002::1 64
	VLAN ID: 4
	Cost: 10

 Copyright © Huawei Technologies Co., Ltd.

Prerequisite

Before you configure OSPFv3 auto FRR, complete the following tasks:

- Configuring link layer protocol parameters and interface IP addresses to ensure that the state of the link layer protocol on the interfaces is Up
- Enabling basic OSPFv3 functions on routers
- Configuring basic BFD for OSPFv3 functions on all devices in area 0 to quickly detect faults

Procedure

Step 1 Configure the MA5800.

1. Configure the IPv6 addresses of the Layer 3 interfaces on the MA5800.

```
huawei(config)#vlan 2 smart
huawei(config)#port vlan 2 0/9 0
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ipv6 enable
huawei(config-if-vlanif2)#ipv6 address 2001:db8:1000::1 64
huawei(config-if-vlanif2)#quit
huawei(config)#vlan 3 smart
huawei(config)#port vlan 3 0/9 1
huawei(config)#interface vlanif 3
huawei(config-if-vlanif3)#ipv6 enable
huawei(config-if-vlanif3)#ipv6 address 2001:db8:1001::1 64
huawei(config-if-vlanif3)#quit
huawei(config)#vlan 4 smart
huawei(config)#port vlan 4 0/9 2
huawei(config)#interface vlanif 4
huawei(config-if-vlanif4)#ipv6 enable
huawei(config-if-vlanif4)#ipv6 address 2001:db8:1002::1 64
huawei(config-if-vlanif4)#quit
```

2. Configure the costs so that the traffic preferentially goes through link T.

```
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ospfv3 cost 10
huawei(config-if-vlanif2)#quit
huawei(config)#interface vlanif 3
huawei(config-if-vlanif3)#ospfv3 cost 15
huawei(config-if-vlanif3)#quit
huawei(config)#interface vlanif 4
huawei(config-if-vlanif4)#ospfv3 cost 10
huawei(config-if-vlanif4)#quit
```

3. Block IP FRR on the specified interface on the MA5800.

```
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ospfv3 frr block
huawei(config-if-vlanif2)#quit
```

4. Enable OSPFv3 auto FRR on the MA5800.

```
huawei(config)#ospfv3
huawei(config-ospfv3-1)#frr
huawei(config-ospfv3-1-frr)#loop-free-alternate
huawei(config-ospfv3-1-frr)#quit
huawei(config-ospfv3-1)#quit
```

5. Save the data.

```
huawei(config)#save
```

----End

 Copyright © Huawei Technologies Co., Ltd.

Result

Run the **display ospfv3 routing** command to display the information about primary/secondary routes configured.

Configuration File

```
vlan 2 smart
port vlan 2 0/9 0
interface vlanif 2
ipv6 enable
ipv6 address 2001:db8:1000::1 64
quit
vlan 3 smart
port vlan 3 0/9 1
interface vlanif 3
ipv6 enable
ipv6 address 2001:db8:1001::1 64
quit
vlan 4 smart
port vlan 4 0/9 2
interface vlanif 4
ipv6 enable
ipv6 address 2001:db8:1002::1 64
quit
interface vlanif 2
ospfv3 cost 10
quit
interface vlanif 3
ospfv3 cost 15
quit
interface vlanif 4
ospfv3 cost 10
quit
interface vlanif 2
ospfv3 frr block
quit
ospfv3
frr
loop-free-alternate
quit
quit
save
```

11.12.7.5 OSPFv3 Reference Standards and Protocols

The following table lists the references.

Document NO.	Description
RFC 2740	This document describes the modifications to OSPF to support IPv6.
RFC 5340	This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

 Copyright © Huawei Technologies Co., Ltd.

Document NO.	Description
draft-ietf-ospf-ospfv3-graceful-restart	This document describes the OSPFv3 graceful restart. The OSPFv3 graceful restart is identical to OSPFv2 except for the differences described in this document. These differences include the format of the grace Link State Advertisements (LSA) and other considerations.
draft-ietf-ospf-ospfv3-mib-11	This memo defines a portion of the Management Information Base (MIB) for use with network management protocols in IPv6-based internets. In particular, it defines objects for managing the Open Shortest Path First Routing Protocol for IPv6.

11.12.8 BGP

11.12.8.1 Introduction

Definition

Border Gateway Protocol (BGP) is a dynamic routing protocol used between Autonomous Systems (ASs). BGP is widely used by Internet Service Providers (ISPs).

As three earlier-released versions of BGP, BGP-1 (defined in RFC 1105), BGP-2 (defined in RFC 1163), and BGP-3 (defined in RFC 1267) are used to exchange reachable inter-AS routes, establish inter-AS paths, avoid routing loops, and apply routing policies between ASs.

Currently, BGP-4 defined in RFC 4271 is used.

As an exterior routing protocol on the Internet, BGP is widely used by Internet service providers (ISPs).

BGP has the following characteristics:

- Unlike an Interior Gateway Protocol (IGP), such as Open Shortest Path First (OSPF) and Routing Information Protocol (RIP), BGP is an Exterior Gateway Protocol (EGP) which controls route advertisement and selects optimal routes between ASs rather than discovering or calculating routes.
- BGP uses Transport Control Protocol (TCP) with port number 179 as the transport layer protocol, which enhances BGP reliability.
  - BGP selects inter-AS routes, which poses high requirements on stability. Therefore, using TCP enhances BGP's stability.
  - BGP peers must be logically connected through TCP. The destination port number is 179 and the local port number is a random value.
- BGP supports Classless Inter-Domain Routing (CIDR).
- When routes are updated, BGP transmits only the updated routes, which reduces bandwidth consumption during BGP route distribution. Therefore,

 Copyright © Huawei Technologies Co., Ltd.

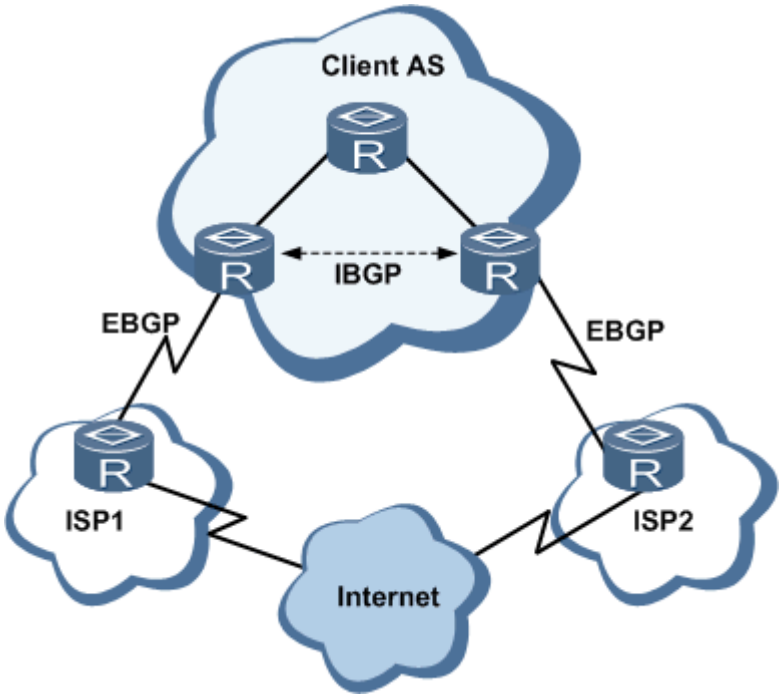
BGP is applicable to the Internet where a large number of routes are transmitted.

- BGP is a distance-vector routing protocol.
- BGP is designed to prevent loops.
  - Between ASs: BGP routes carry information about the ASs along the path. The routes that carry the local AS number are discarded to prevent inter-AS loops.
  - Within an AS: BGP does not advertise routes learned in an AS to BGP peers in the AS to prevent intra-AS loops.
- BGP provides many routing policies to flexibly select and filter routes.
- BGP provides a mechanism that prevents route flapping, which effectively enhances Internet stability.
- BGP can be easily extended.

Purpose

BGP transmits route information between ASs. It, however, is not required in all scenarios.

Figure 11-151 BGP networking



BGP is required in the following scenarios:

- On the network shown in **Figure 11-151**, users need to be connected to two or more ISPs. The ISPs need to provide all or part of the Internet routes for the users. Routers, therefore, need to select the optimal route through the AS of an ISP to the destination based on the attributes carried in BGP routes.

 Copyright © Huawei Technologies Co., Ltd.

- The AS_Path attribute needs to be transmitted between users in different organizations.
- Users need to transmit VPN routes through a Layer 3 VPN.

BGP is not required in the following scenarios:

- Users are connected to only one ISP.
- The ISP does not need to provide Internet routes for users.
- ASs are connected through default routes.

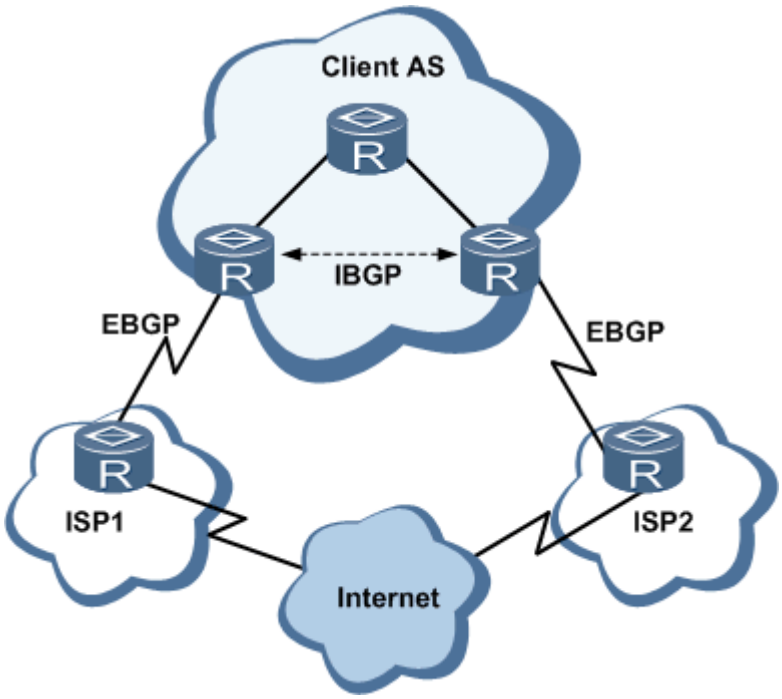
11.12.8.2 Principles

11.12.8.2.1 Basic Principle

BGP Operating Modes

BGP is called Internal BGP (IBGP) when it runs within an AS; it is called External BGP (EBGP) when it runs between ASs, as shown in **Figure 11-152**.

Figure 11-152 BGP operating modes



Roles in Transmitting BGP Messages

- Speaker: Any Router that sends BGP messages is called a BGP speaker. The speaker receives or generates new routing information and then advertises the routing information to other BGP speakers. After receiving a route from another AS, a BGP speaker compares the route with its local routes. If the route is better than its local routes, or the route is new, the speaker advertises this route to all other BGP speakers.

 Copyright © Huawei Technologies Co., Ltd.

- Peer: BGP speakers that exchange messages with each other are called peers.

BGP Messages

BGP runs by sending five types of messages: Open, Update, Notification, Keepalive, and Route-refresh.

- Open: The first message sent after a TCP connection is set up is an Open message, which is used to set up BGP peer relationships. After a peer receives an Open message and the peer negotiation is successful, the peer sends a Keepalive message to confirm and maintain the peer relationship. Then, peers can exchange Update, Notification, Keepalive, and Route-refresh messages.
- Update: This type of message is used to exchange routes between BGP peers.
  - An Update message can advertise multiple reachable routes with the same attributes. These route attributes are applicable to all destination addresses (expressed by IP prefixes) in the Network Layer Reachability Information (NLRI) field of the Update message.
  - An Update message can be used to delete multiple unreachable routes. Each route is identified by its destination address (using the IP prefix), which identifies the routes previously advertised between BGP speakers.
  - An Update message can be used only to delete routes. In this case, it does not need to carry the route attributes or NLRI. In addition, an Update message can be used only to advertise reachable routes. In this case, it does not need to carry information about the deleted routes.
- Notification: When BGP detects an error, it sends a Notification message to its peer. The BGP connection is then torn down immediately.
- Keepalive: BGP periodically sends Keepalive messages to peers to maintain peer relationships.
- Route-refresh: This type of message is used to request that the peer resend all reachable routes.

  If all BGP routers are enabled with the Route-refresh capability and the import policy of BGP changes, the local BGP router sends a Route-refresh message to its peers. After receiving the Route-refresh message, the peers resend their routing information to the local BGP router. In this manner, BGP routing tables are dynamically refreshed and new routing policies are used without tearing down BGP connections.

BGP Finite State Machine

The BGP Finite State Machine (FSM) has six states: Idle, Connect, Active, OpenSent, OpenConfirm, and Established.

Three common states during the establishment of BGP peer relationships are Idle, Active, and Established.

- In the Idle state, BGP denies all connection requests. This is the initial status of BGP.
- In the Connect state, BGP decides subsequent operations after a TCP connection is established.
- In the Active state, BGP attempts to establish a TCP connection.

 Copyright © Huawei Technologies Co., Ltd.

- In the OpenSent state, BGP is waiting for an Open message from the peer.
- In the OpenConfirm state, BGP is waiting for a Notification or Keepalive message.
- In the Established state, BGP peers can exchange Update, Route-refresh, Keepalive, and Notification messages.

The BGP peer relationship can be established only when both BGP peers are in the Established state. Both peers send Update messages to exchange routes.

BGP Processing

- BGP adopts TCP as its transport layer protocol. Therefore, a TCP connection must be available between the peers. BGP peers negotiate parameters by exchanging Open messages to establish a BGP peer relationship.
- After the peer relationship is established, BGP peers exchange BGP routing tables. BGP does not periodically update a routing table. When BGP routes change, BGP updates the changed BGP routes in the BGP routing table by sending Update messages.
- BGP sends Keepalive messages to maintain the BGP connection between peers.
- After detecting an error on a network, BGP sends a Notification message to report the error and the BGP connection is torn down.

BGP Attributes

BGP route attributes are a set of parameters that describe specific BGP routes. With BGP route attributes, BGP can filter and select routes. BGP route attributes are classified into the following types:

- Well-known mandatory: This type of attribute can be identified by all BGP routers and must be carried in Update messages. Without this attribute, errors occur in the routing information.
- Well-known discretionary: This type of attribute can be identified by all BGP routers. This type of attribute is optional and, therefore, is not necessarily carried in Update messages.
- Optional transitive: This indicates the transitive attribute between ASs. A BGP router may not recognize this attribute, but the router still receives it and advertises it to other peers.
- Optional non-transitive: If a BGP router does not recognize this type of attribute, the router does not advertise it to other peers.

The most common BGP route attributes are as follows:

- Origin

  The Origin attribute defines the origin of a route. The Origin attribute is classified into the following types:

  – Interior Gateway Protocol (IGP): This attribute type has the highest priority. IGP is the Origin attribute for routes obtained through an IGP of the AS from which the routes originate. For example, the Origin attribute of the routes imported to the BGP routing table using the **network** command is IGP.

 Copyright © Huawei Technologies Co., Ltd.

  - Exterior Gateway Protocol (EGP): This attribute type has the second highest priority. The Origin attribute of the routes obtained through EGP is EGP.
  - Incomplete: This attribute type has the lowest priority. Incomplete is the Origin attribute type of all routes that do not have the IGP or EGP Origin attribute. For example, for the routes imported using the **import-route** command by BGP, the Origin attribute is Incomplete.

- AS_Path

  The AS-Path attribute records all ASs through which a route passes from the local end to the destination in distance-vector (DV) order.

  When a BGP speaker advertises a local route:

  - When advertising the route beyond the local AS, the BGP speaker adds the local AS number to the AS_Path list and then advertises it to the neighboring Routers through Update messages.
  - When advertising the route within the local AS, the BGP speaker creates an empty AS_Path list in an Update message.

  When a BGP speaker advertises a route learned from the Update messages of another BGP speaker:

  - When advertising the route beyond the local AS, the BGP speaker adds the local AS number to the left of the AS_Path list. From the AS_Path attribute, the BGP router that receives the route learns the ASs through which the route passes to the destination. The number of the AS that is nearest to the local AS is placed on the left of the list, while other AS numbers are listed in sequence.
  - When advertising the route within the local AS, the BGP speaker does not change the AS_Path attribute.

  The AS_Path attribute has four types:

  - AS_Sequence: a sequenced set of numbers of the ASs that a route passes through from a local end to the destination
  - AS_Set: an unsequenced set of numbers of the ASs that a route passes through from a local end to the destination The AS_Set attribute is used in the route aggregation scenario. After route aggregation, the device cannot sequence the numbers of ASs that specific routes pass through, so the AS_Set attribute is used to record the unsequenced AS numbers. No matter how many AS numbers an AS_Set contains, BGP regards the AS_Set as one AS number to calculate routes.
  - AS_Confed_Sequence: a sequenced set of sub-AS numbers in a confederation
  - AS_Confed_Set: an unsequenced set of sub-AS numbers in a confederation. The AS_Confed_Set attribute is used in the route aggregation scenario in a confederation.

  The AS_Confed_Sequence and AS_Confed_Set attributes are used to prevent route loops and to select routes among the various sub-ASs in a confederation.

- Next_Hop

  The Next_Hop attribute in BGP is different from that in IGP. It is not necessarily the IP address of a neighboring Router. Generally, the Next_Hop attribute complies with the following rules:

 Copyright © Huawei Technologies Co., Ltd.

  - When advertising a route to an EBGP peer, a BGP speaker sets the next hop of the route to the address of the local interface through which the BGP peer relationship is established.
  - When advertising a local route to an IBGP peer, a BGP speaker sets the next hop of the route to the address of the local interface through which the BGP peer relationship is established.
  - When advertising a route learned from an EBGP peer to an IBGP peer, the BGP speaker does not change the next hop of the route.
- MED

  The Multi-Exit-Discriminator (MED) is transmitted only between two neighboring ASs. The AS that receives the MED does not advertise it to a third AS.

  Similar to the metric used by an IGP, the MED is used to determine the optimal route when traffic enters an AS. When a BGP peer obtains multiple routes to the same destination address but with different next hops through EBGP peers, the route with the smallest MED value is selected as the optimal route if all other attributes are the same.
- Local_Pref

  The Local_Pref attribute indicates the BGP priority of a route. It is available only to IBGP peers and is not advertised to other ASs.

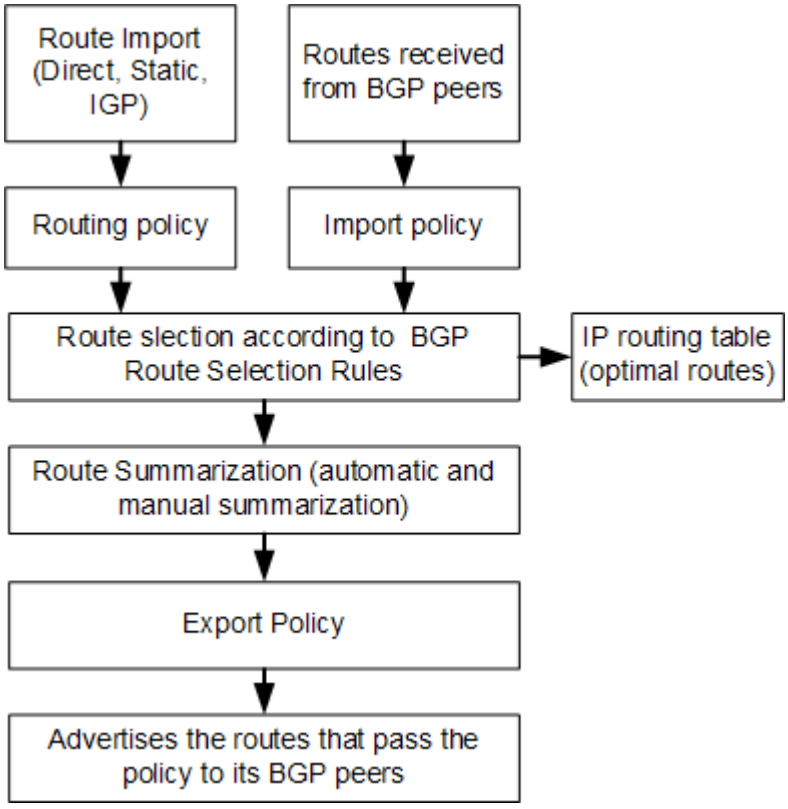
  The Local_Pref attribute is used to determine the optimal route when traffic leaves an AS. When a BGP Router obtains multiple routes to the same destination address but with different next hops through IBGP peers, the route with the largest Local_Pref value is selected.

11.12.8.2.2 BGP Route Processing

Figure 11-153 shows how BGP processes routes. BGP routes can be imported from other protocols or learned from peers. To reduce the routing size, you can configure route summarization after BGP selects routes. In addition, you can configure route-policies and apply them to route import, receipt, or advertisement to filter routes or modify route attributes.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-153 BGP route processing



Route Import

BGP itself cannot discover routes. Therefore, it needs to import other protocol routes, such as IGP routes or static routes, to the BGP routing table. Imported routes can be transmitted within an AS or between ASs.

BGP routes are imported in either of the following modes:

- The **import** command imports routes based on protocol types, such as RIP routes, OSPF routes, Intermediate System to Intermediate System (IS-IS) routes, static routes, or direct routes.
- The **network** command imports a route with the specified prefix and mask to the BGP routing table, which is more precise than the previous mode.

BGP Route Selection

When multiple routes to the same destination are available, BGP selects routes based on the following rules:

1. Prefers routes in descending order of Valid, Not Found, and Invalid after BGP origin AS validation states are applied to route selection in a scenario where the device is connected to an RPKI server.
2. Prefers the route with the largest PreVal value.

   PrefVal is Huawei-specific. It is valid only on the device where it is configured.
3. Prefers the route with the highest Local_Pref.

   If a route does not carry Local_Pref, the default value 100 takes effect. To change the value, run the **default local-preference** command.

 Copyright © Huawei Technologies Co., Ltd.

4. Prefers a locally originated route to a route learned from a peer.

   Locally originated routes include routes imported using the **network** or **import-route** command, as well as manually and automatically summarized routes.

   a. Prefers a summarized route over a non-summarized route.

   b. Prefers a route obtained using the **aggregate** command over a route obtained using the **summary automatic** command.

   c. Prefers a route imported using the **network** command over a route imported using the **import-route** command.

5. Prefers the route with the shortest AS_Path length.

   - The AS_CONFED_SEQUENCE and AS_CONFED_SET are not included in the AS_Path length.
   - During route selection, a router assumes that an AS_SET carries only one AS number regardless of the actual number of ASs it is carrying.
   - If the **bestroute as-path-ignore** command is run, BGP no longer compares the AS_Path attribute.

6. Prefers the route with the Origin type as IGP, EGP, and Incomplete in descending order.

7. Prefers the route with the smallest MED value.

   - BGP compares the MEDs of only routes from the same AS (excluding confederation sub-ASs). MEDs of two routes are compared only when the first AS number in the AS_Sequence (excluding AS_Confed_Sequence) of one route is the same as its counterpart in the other route.
   - If a route does not carry MED, BGP considers its MED as the default value (0) during route selection. If the **bestroute med-none-as-maximum** command is run, BGP considers its MED as the largest MED value (4294967295).
   - If the **compare-different-as-med** command is run, BGP compares MEDs of routes even when the routes are received from peers in different ASs. Do not run this command unless the ASs use the same IGP and route selection mode. Otherwise, a loop may occur.
   - If the **deterministic-med** command is run, routes are no longer selected in the sequence in which they are received.

8. Prefers EBGP routes to IBGP routes.

   EBGP, IBGP, LocalCross, and RemoteCross routes are preferred in descending order.

   If the ERT of a VPNv4 route in the routing table of a VPN instance on a PE matches the IRT of another VPN instance on the PE, the VPNv4 route is added to the routing table of the second VPN instance. This route is called a LocalCross route. If the ERT of a VPNv4 route learned from a remote PE matches the IRT of a VPN instance on the local PE, the VPNv4 route is added to the routing table of that VPN instance. This route is called a RemoteCross route.

9. Prefers the route that is iterated to an IGP route with the smallest cost.

   If the **bestroute igp-metric-ignore** command is run, BGP no longer compares the IGP cost.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

With load balancing, if the preceding conditions are equal and multiple external routes with the same AS_Path are available, load balancing is performed among them. The number of routes load-balancing traffic must be less than or equal to the configured number. After the **load-balancing as-path-ignore** command is run, the routes with different As_Path values can load-balance traffic.

10. Prefers the route with the shortest Cluster_List length.
11. Prefers the route advertised by the Router with the smallest router ID.

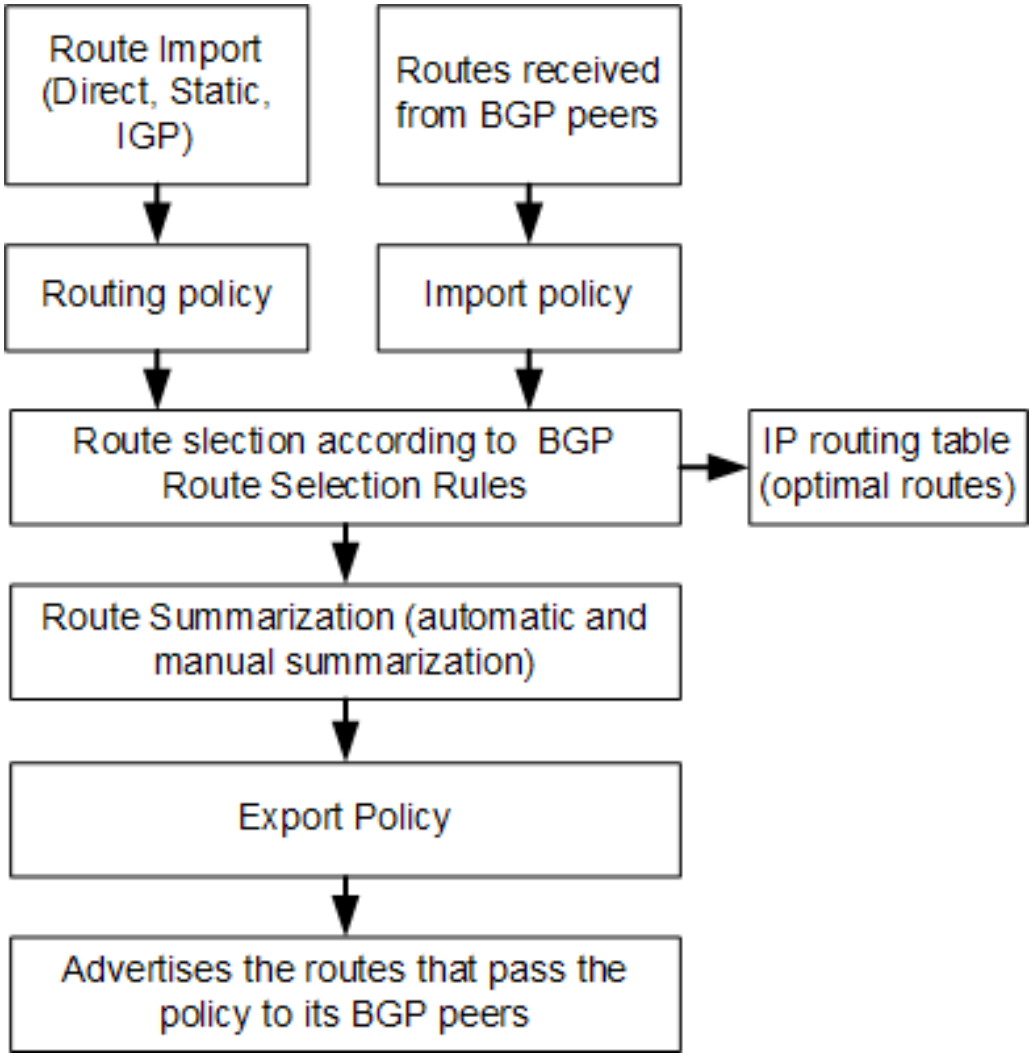
    If the **bestroute router-id-ignore** command is run, router IDs do not determine which route is selected for BGP.

    NOTE

    If each route carries an Originator_ID, the originator IDs rather than router IDs are compared during route selection. The route with the smallest Originator_ID is preferred.
12. Prefers the route learned from the peer with the smallest IP address.

For details about the BGP route selection process, see **Figure 11-154**.

Figure 11-154 BGP route selection process



 Copyright © Huawei Technologies Co., Ltd.

Route Summarization

On a large-scale network, the BGP routing table can be very large. Route summarization can reduce the size of the routing table.

Route summarization is the process of summarizing specific routes with the same IP prefix into a summarized route. After route summarization, BGP advertises only the summarized route rather than all specific routes to BGP peers.

BGP supports automatic and manual route summarization.

- Automatic route summarization: takes effect on the routes imported by BGP. With automatic route summarization, the specific routes for the summarization are suppressed, and BGP summarizes routes based on the natural network segment and sends only the summarized route to BGP peers. For example, 10.1.1.1/24 and 10.2.1.1/24 are summarized into 10.0.0.0/8, which is a Class A address.
- Manual route summarization: takes effect on routes in the local BGP routing table. With manual route summarization, users can control the attributes of the summarized route and determine whether to advertise the specific routes.

IPv4 supports both automatic and manual route summarization, while IPv6 supports only manual route summarization.

BGP Route Advertisement

BGP adopts the following policies to advertise routes:

- When there are multiple valid routes, a BGP speaker advertises only the optimal route to its peers.
- A BGP speaker advertises the routes learned from EBGP peers to all BGP peers, including EBGP peers and IBGP peers.
- A BGP speaker does not advertise the routes learned from an IBGP peer to other IBGP peers.
- A BGP speaker advertises the routes learned from IBGP peers to its EBGP peers.
- A BGP speaker advertises all BGP optimal routes to new peers after peer relationships are established.

11.12.8.2.3 Peer Group

A peer group is a set of peers with the same policies. When a peer joins a peer group, it inherits the configurations of the peer group. If the configurations of the peer group change, the configurations of all the peers in the group changes accordingly.

On a large-scale BGP network, there are many peers and most of them need the same policies. Therefore, some commands need to be run repeatedly on each peer. Configuring a peer group can simplify the configuration.

Each peer in a peer group can be configured with unique policies to advertise and receive routes.

 Copyright © Huawei Technologies Co., Ltd.

11.12.8.2.4 Route Dampening

Route instability is reflected by route flapping. When a route flaps, it repeatedly disappears from the routing table and then reappears.

NOTE

A route is added to the routing table, and then is withdrawn. This process is called one route flapping.

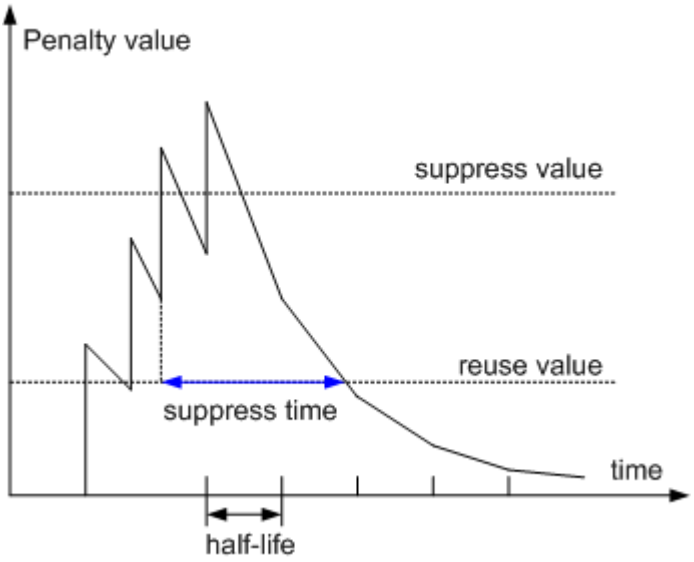
If route flapping occurs, a Router sends an Update packet to its peers. After the peers receive the Update packet, they recalculate routes and update their routing tables. Frequent route flapping consumes lots of bandwidth and CPU resources and can even affect network operations.

Route dampening can address this problem. In most cases, BGP is deployed on complex networks where routes change frequently. To reduce the impact of frequent route flapping, BGP adopts route dampening to suppress unstable routes.

BGP dampening measures route stability using a penalty value. The greater the penalty value, the more unstable a route. Each time route flapping occurs, BGP adds a penalty value to this route. If a route changes from active to inactive, the penalty value increases by 1000. If a route is updated when it is active, the penalty value increases by 500. When the penalty value of a route exceeds the suppress value, the route is suppressed. As a result, BGP does not add the route to the routing table or advertise any Update message to BGP peers.

The penalty value of a suppressed route reduces by half after a half-life period. When the penalty value decreases to the reuse value, the route is reusable and is added to the routing table. At the same time, BGP advertises an Update message to BGP peers. The penalty value, suppression threshold, and half-life are configurable. **Figure 11-155** shows the process of BGP route dampening.

Figure 11-155 BGP route dampening



Route dampening applies only to EBGP routes. IBGP routes cannot be dampened because IBGP routing tables contain the routes from the local AS, which require

 Copyright © Huawei Technologies Co., Ltd.

that the forwarding entries be the same on IBGP peers in the AS. If IBGP routes are dampened, the forwarding entries may be inconsistent because dampening parameters may vary among these IBGP peers.

11.12.8.2.5 Community Attribute

A community is a set of destination addresses with the same characteristics. It is four bytes long. On the , the community attribute is expressed in the format of aa:nn or as a community number.

- aa:nn: aa indicates an AS number and nn indicates the community identifier defined by an administrator. The value of aa or nn ranges from 0 to 65535, which is configurable. For example, if a route is from AS 100 and the community identifier defined by the administrator is 1, the community is 100:1.
- Community number: It is an integer ranging from 0 to 4294967295. As defined in RFC 1997, numbers from 0 (0x00000000) to 65535 (0x0000FFFF) and from 4294901760 (0xFFFF0000) to 4294967295 (0xFFFFFFFF) are reserved.

The community attribute is used to simplify the application, maintenance, and management of routing policies. With the community attribute, a group of BGP peers in multiple ASs can share the same routing policy. The community attribute is a route attribute. It is transmitted between BGP peers and is not restricted by the AS. Before advertising a route with the community attribute to peers, a BGP peer can change the original community attribute of this route.

The peers in a peer group share the same policy, while the routes with the same community attribute share the same policy.

The well-known communities are described in the next section. Users can also create their own communities to filter routes.

Well-known Community

Table 11-79 lists well-known communities of BGP routes.

Table 11-79 Well-known communities of BGP routes

Community Name	Community Identifier	Description
Internet	0 (0x00000000)	By default, all routes belong to the Internet community. A route with this attribute can be advertised to all BGP peers.
No_Export	4294967041 (0xFFFFFF01)	A route with this attribute cannot be advertised beyond the local AS.
No_Advertise	4294967042 (0xFFFFFF02)	A route with this attribute cannot be advertised to any other BGP peers.

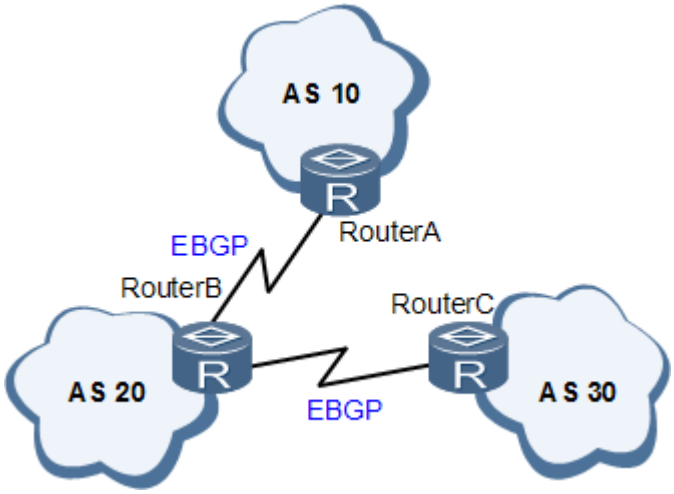
 Copyright © Huawei Technologies Co., Ltd.

Community Name	Community Identifier	Description
No_Export_Subconfed	4294967043 (0xFFFFFF03)	A route with this attribute cannot be advertised beyond the local AS or to other sub-ASs.

Usage Scenario

On the network shown in **Figure 11-156**, EBGP connections are established between Router A and Router B, and between Router B and Router C. If the community attribute of No_Export is configured on Router A and Router A sends a route with the community attribute to Router B, Router B does not advertise the route to other ASs after receiving it.

Figure 11-156 Networking for BGP communities



11.12.8.2.6 Route Reflector

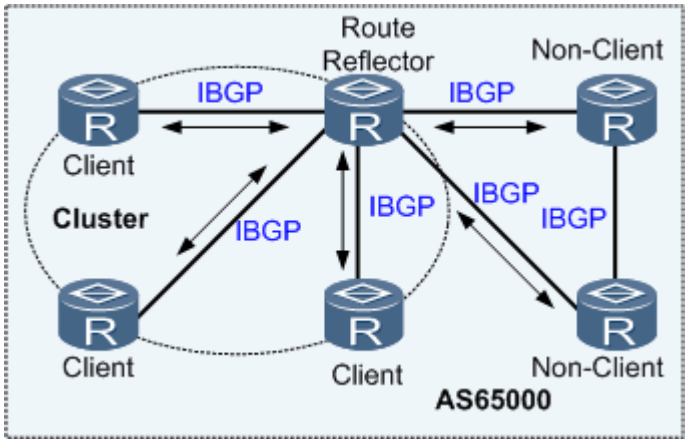
Fully meshed connections need to be established between IBGP peers to ensure the connectivity between IBGP peers. If there are n Routers in an AS, n x (n-1)/2 IBGP connections need to be established. When there are a lot of IBGP peers, network resources and CPU resources are greatly consumed. Route reflection can solve the problem.

In an AS, one Router functions as a Route Reflector (RR) and the other Routers as clients. The clients establish IBGP connections with the RR. The RR and its clients form a cluster. The RR reflects routes among clients, and BGP connections do not need to be established between the clients.

A BGP peer that functions as neither an RR nor a client is called a non-client. A non-client must establish full meshed connections with the RR and all the other non-clients, as shown in **Figure 11-157**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-157 Networking with an RR



Applications

After an RR receives routes from its peers, it selects the optimal route based on BGP route selection policies and performs one of the following operations:

- If the optimal route is from a non-client IBGP peer, the RR advertises the route to all clients.
- If the optimal route is from a client, the RR advertises the route to all non-clients and clients.
- If the optimal route is from an EBGP peer, the RR advertises the route to all clients and non-clients.

An RR is easy to configure because it only needs to be configured on the RR itself, and clients do not need to know whether they are clients.

On some networks, if fully meshed connections have already been established among clients of an RR, they can exchange routing information directly. In this case, route reflection among the clients is unnecessary and occupies bandwidth. For example, on the , route reflection can be disabled, but the routes between clients and non-clients are still exchanged. By default, route reflection between clients is enabled.

On the , an RR can change various attributes of BGP routes, such as the AS_Path, MED, Local_Pref, and community attributes.

Originator_ID

Originator_ID and Cluster_List, defined in RFC 2796, are used to detect and prevent routing loops.

The Originator_ID attribute is four bytes long and is generated by an RR. It carries the router ID of the originator of the route in the local AS.

- When a route is reflected by an RR for the first time, the RR adds the Originator_ID to this route. If a route already carries an Originator_ID, the RR does not add a new one.

 Copyright © Huawei Technologies Co., Ltd.

- After receiving the route, a BGP speaker checks whether the Originator_ID is the same as its router ID. If Originator_ID is the same as its router ID, the BGP speaker discards this route.

Cluster_List

To prevent routing loops between ASs, a BGP Router uses the AS_Path attribute to record the ASs through which a route passes. Routes with the local AS number are discarded by the Router. To prevent routing loops within an AS, IBGP peers do not advertise routes learned from the local AS.

With RR, IBGP peers can advertise routes learned from the local AS to each other. However, the Cluster_List attribute must be deployed to prevent routing loops within the AS.

An RR and its clients form a cluster. In an AS, each RR is uniquely identified by a Cluster_ID.

Similar to an AS_Path, a Cluster_List is composed of a series of Cluster_IDs and is generated by an RR. It records all the RRs through which a route passes.

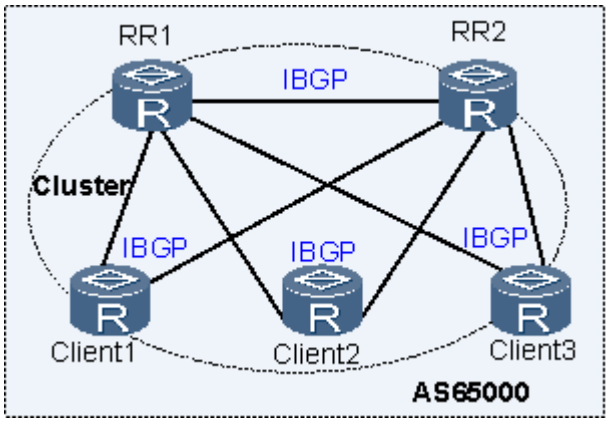
- Before an RR reflects a route between its clients or between its clients and non-clients, the RR adds the local Cluster_ID to the head of the Cluster_List. If a route does not carry any Cluster_List, the RR creates one for the route.
- After the RR receives an updated route, it checks the Cluster_List of the route. If the RR finds that its cluster ID is included in the Cluster_List, the RR discards the route. If its cluster ID is not included in the Cluster_List, the RR adds its cluster ID to the Cluster_List and then reflects the route.

Backup RR

To enhance network reliability and prevent single points of failure, more than one route reflector needs to be configured in a cluster. The route reflectors in the same cluster must share the same Cluster_ID to prevent routing loops.

With backup RRs, clients can receive multiple routes to the same destination from different RRs. The clients then apply route selection policies to choose the optimal route.

Figure 11-158 Backup RR



 Copyright © Huawei Technologies Co., Ltd.

On the network shown in **Figure 11-158**, RR1 and RR2 are in the same cluster. RR1 and RR2 establish an IBGP connection so that each RR is a non-client of the other RR.

- If Client 1 receives an updated route from an external peer, Client 1 advertises the route to RR1 and RR2 through IBGP.
- After receiving the updated route, RR1 reflects the route to other clients (Client 2 and Client 3) and the non-client (RR2) and adds the local Cluster_ID to the head of the Cluster_List.
- After receiving the reflected route, RR2 checks the Cluster_List. RR2 finds that its Cluster_ID is contained in the Cluster_List; therefore, it discards the updated route.

If RR1 and RR2 are configured with different Cluster_IDs, each RR receives both the route from Client 1 and the updated route reflected from the other RR. Therefore, configuring the same Cluster_ID for RR1 and RR2 reduces the number of routes that each RR receives and memory consumption.

NOTE

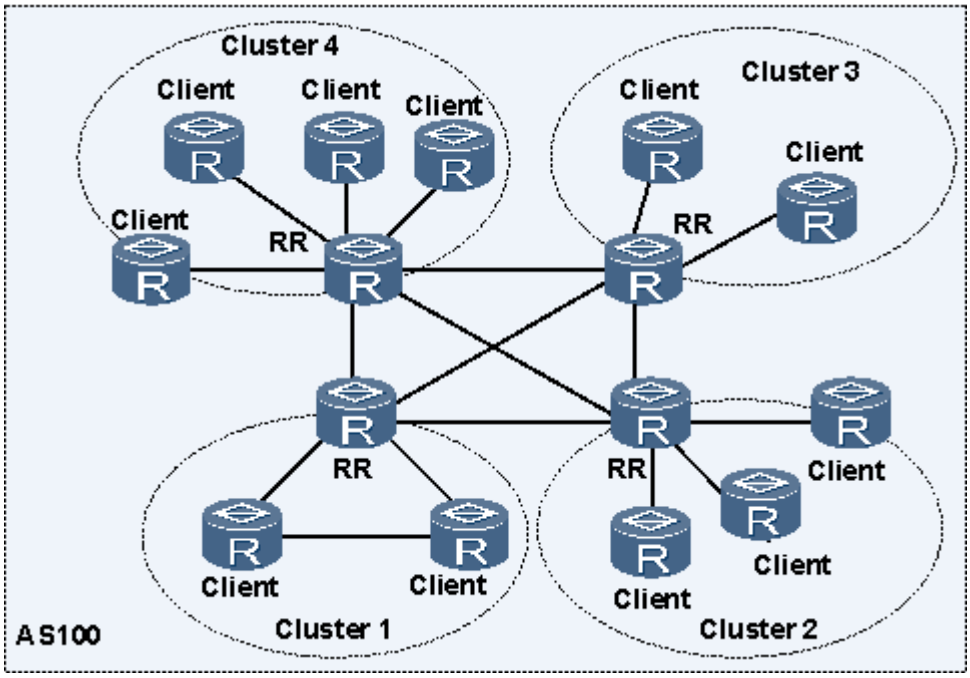
The application of Cluster_List prevents routing loops among RRs in the same AS.

Multiple Clusters in an AS

Multiple clusters may exist in an AS. RRs are IBGP peers of each other. An RR can be configured as a client or non-client of another RR.

For example, the backbone network shown in **Figure 11-159** is divided into multiple clusters. Each RR is configured as a non-client of the other RRs, and these RRs are fully meshed. Each client establishes IBGP connections with only the RR in the same cluster. In this manner, all BGP peers in the AS can receive reflected routes.

Figure 11-159 Multiple clusters in an AS



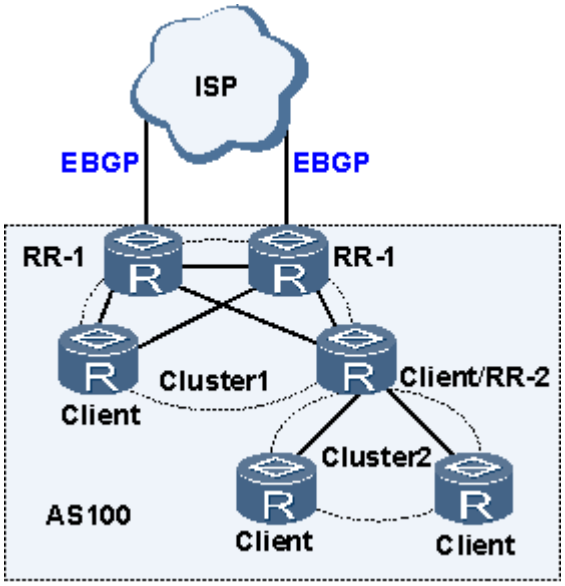
 Copyright © Huawei Technologies Co., Ltd.

Hierarchical Reflector

Hierarchical reflectors are deployed on live networks. On the network shown in **Figure 11-160**, the ISP provides Internet routes for AS 100. Two EBGP connections are established between the ISP and AS 100. AS 100 is divided into two clusters. The four Routers in Cluster 1 are core routers.

- Two Level-1 RRs (RR-1s) are deployed in Cluster 1, which ensures the reliability of the core layer of AS 100. The other two Routers in the core layer are clients of RR-1s.
- One Level-2 RR (RR-2) is deployed in Cluster 2. RR-2 is a client of RR-1.

Figure 11-160 Hierarchical reflector

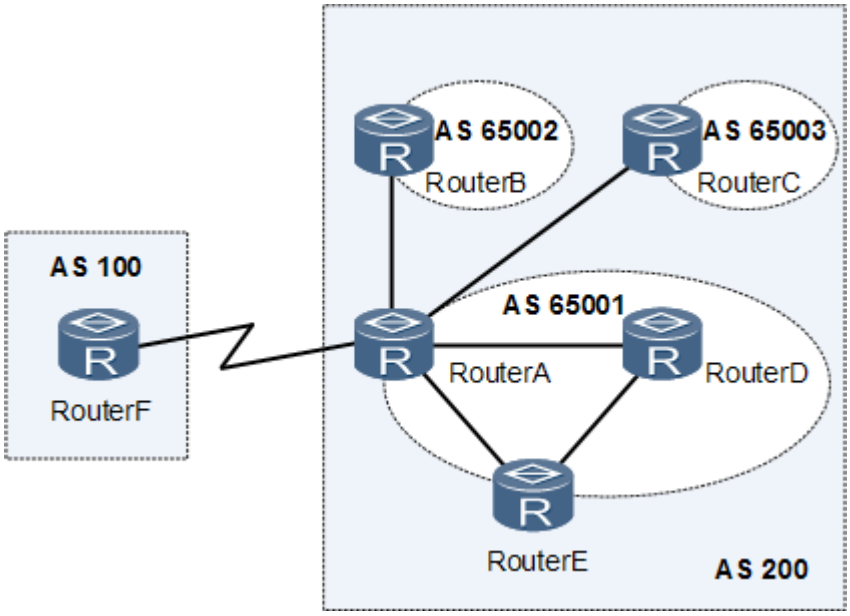


11.12.8.2.7 BGP Confederation

Besides RR, BGP confederation can also reduce IBGP connections in an AS. It divides an AS into several sub-ASs. Fully meshed IBGP connections are established in each sub-AS, and fully meshed EBGP connections are established between sub-ASs. As shown in **Figure 11-161**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-161 BGP confederation



As shown in **Figure 11-161**, there are multiple BGP routers in AS 200. To reduce the number of IBGP connections, AS 200 is divided into three sub-ASs: AS 65001, AS 65002, and AS 65003. In AS 65001, fully meshed IBGP connections are established between the three routers.

BGP speakers outside a confederation such as Router F in AS 100, do not know the existence of the sub-ASs (AS 65001, AS 65002, and AS 65003) in the confederation. The confederation ID is the AS number that is used to identify the entire confederation. For example, AS 200 in **Figure 11-161** is the confederation ID.

Applications and Limitations

The confederation needs to be configured on each Router, and the Router that joins the confederation must support the confederation function.

BGP speakers need to be reconfigured when a network in non-confederation mode switches to confederation mode. As a result, the logical topology changes accordingly.

On large-scale BGP networks, the RR and confederation can both be used.

11.12.8.2.8 MP-BGP

Conventional BGP-4 manages only IPv4 unicast routing information, and inter-AS transmission of packets of other network layer protocols, such as IPv6 and multicast, is limited.

To support multiple network layer protocols, the Internet Engineering Task Force (IETF) extends BGP-4 to Multiprotocol Extensions for BGP-4 (MP-BGP). RFC 4760 defines the MP-BGP standard. MP-BGP is forward compatible. Specifically, Routers supporting MP-BGP can communicate with the Routers that do not support MP-BGP.

 Copyright © Huawei Technologies Co., Ltd.

Extended Attributes

BGP-4 Update packets carry three IPv4-related attributes: NLRI (Network Layer Reachable Information), Next_Hop, and Aggregator. Aggregator contains the IP address of the BGP speaker that performs route aggregation.

To carry information about multiple network layer protocols in NLRI and Next_Hop, MP-BGP introduces the following route attributes:

- MP_REACH_NLRI: indicates the multiprotocol reachable NLRI. It is used to advertise a reachable route and its next hop.
- MP_UNREACH_NLRI: indicates the multiprotocol unreachable NLRI. It is used to delete an unreachable route.

The preceding two attributes are optional non-transitive. Therefore, the BGP speakers that do not support MP-BGP will ignore the information carried in the two attributes and do not advertise the information to other peers.

Address Family

The Address Family Information field consists of a 2-byte Address Family Identifier (AFI) and a 1-byte Subsequent Address Family Identifier (SAFI).

BGP uses address families to distinguish different network layer protocols. For the values of address families, see RFC 3232 (Assigned Numbers). The supports multiple MP-BGP extension applications, such as VPN extension and IPv6 extension, which are configured in their respective address family views.

11.12.8.2.9 BGP GR

Graceful restart (GR) is one of the high availability (HA) technologies, which comprise a series of comprehensive technologies such as fault-tolerant redundancy, link protection, faulty node recovery, and traffic engineering. As a fault-tolerant redundancy technology, GR ensures normal forwarding of data during the restart of routing protocols to prevent interruption of key services. Currently, GR has been widely applied to the master/slave switchover and system upgrade.

GR is usually used when the active route processor (RP) fails because of a software or hardware error, or used by an administrator to perform the master/slave switchover.

Prerequisite for Implementation

On a traditional routing device, a processor implements both control and forwarding. The processor finds routes based on routing protocols, and maintains the routing table and forwarding table of the device. Mid-range and high-end devices generally adopt the multi-RP structure to improve forwarding performance and reliability. The processor in charge of routing protocols is located on the main control board, whereas the processor responsible for data forwarding is located on the interface board. The design helps to ensure the continuity of packet forwarding on the interface board during the restart of the main processor. The technology that separates control from forwarding satisfies the prerequisite for GR implementation.

 Copyright © Huawei Technologies Co., Ltd.

At present, a GR-capable device must have two main control boards. In addition, the interface board must have an independent processor and memory.

Related Concepts

The concepts related to GR are as follows:

- GR Restarter: indicates a device that performs master/slave switchover triggered by the administrator or a failure. A GR Restarter must support GR.
- GR helper: indicates the neighbor of a GR Restarter. A GR helper must support GR.
- GR session: indicates a session, through which a GR Restarter and a GR helper can negotiate GR capabilities.
- GR time: indicates the time when the GR helper finds that the GR Restarter is Down but keeps the topology information or routes obtained from the GR Restarter.
- End-of-RIB (EOR): indicates a BGP information, notifying a peer BGP that the first route upgrade is finished after the negotiation.
- EOR timer: indicates a maximum time of a local device waiting for the EOR information sent from the peer. If the local device does not receive the EOR information from the peer within the EOR timer, the local device will select an optimal route from the current routes.

Principles

Principles of BGP GR are as follows:

1. During BGP peer relationship establishment, devices negotiate GR capabilities by sending supported GR capabilities to each other.
2. When detecting the master/slave switchover of the GR Restarter, a GR helper does not delete the routing information and forwarding entries related to the GR Restarter within the GR time, but waits to re-establish a BGP connection with the GR Restarter.
3. After the master/slave switchover, the GR Restarter receives routes from all the negotiated peers with GR capabilities before the switchover, and starts the EOR timer. The GR Restarter selects a route when either of the following conditions is met:
   - The GR Restarter receives the EOR information of all peers and the EOR timer is deleted.
   - The EOR timer times out but the GR Restarter receives no EOR information from all peers.
4. The GR Restarter sends the optimal route to the GR Helper and the GR Helper starts the EOR timer. The GR Helper quits GR when either of the following conditions is met:
   - The GR Helper receives the EOR information from the GR Restarter and the EOR timer is deleted.
   - The EOR timer times out and the GR Helper receives no EOR information from the GR Restarter.

   NOTE

   Currently, the system supports only the GR helper.

 Copyright © Huawei Technologies Co., Ltd.

GR Reset

Currently, BGP does not support dynamic capability negotiation. Therefore, each time a new BGP capability (such as the IPv4, IPv6, VPNv4, and VPNv6 capabilities) is enabled on a BGP speaker, the BGP speaker tears down existing sessions with its peer and renegotiates BGP capabilities. This process will interrupt ongoing services.

To prevent the service interruptions, the provides the GR reset function that enables the to reset a BGP session in GR mode. With the GR reset function configured, when you enable a new BGP capability on the BGP speaker, the BGP speaker enters the GR state, resets the BGP session, and renegotiates BGP capabilities with the peer. In the whole process, the BGP speaker re-establishes the existing sessions but does not delete the routing entries for the existing sessions, so that the existing services are not interrupted.

Benefits

Through BGP GR, the forwarding is not interrupted. In addition, the flapping of BGP occurs only on the neighbors of the GR Restarter, and does not occur in the entire routing domain. This is important for BGP that needs to process a large number of routes.

11.12.8.2.10 BGP Dynamic Update Peer-Groups

As the routing table increases in size and the network topology increases in complexity, BGP needs to be able to support more peers. When the Router needs to send a large number of routes to many BGP peers and most of the peers share the same outbound policy, if the Router groups each route and then send the route to each peer, the efficiency is low.

The dynamic update peer-groups feature can address this problem. It considers all the BGP peers with the same outbound policy as an update-group. With the dynamic update peer-groups feature, each route to be sent is grouped once for all and then sent to all peers in the update-group, improving grouping efficiency and forwarding performance exponentially.

Usage Scenario

The BGP dynamic update peer-groups feature is applicable to the following scenarios:

- Scenario with an international gateway
- Scenario with an RR
- Scenario where routes received from EBGP neighbors need to be sent to all IBGP peers

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-162 Networking for the international gateway

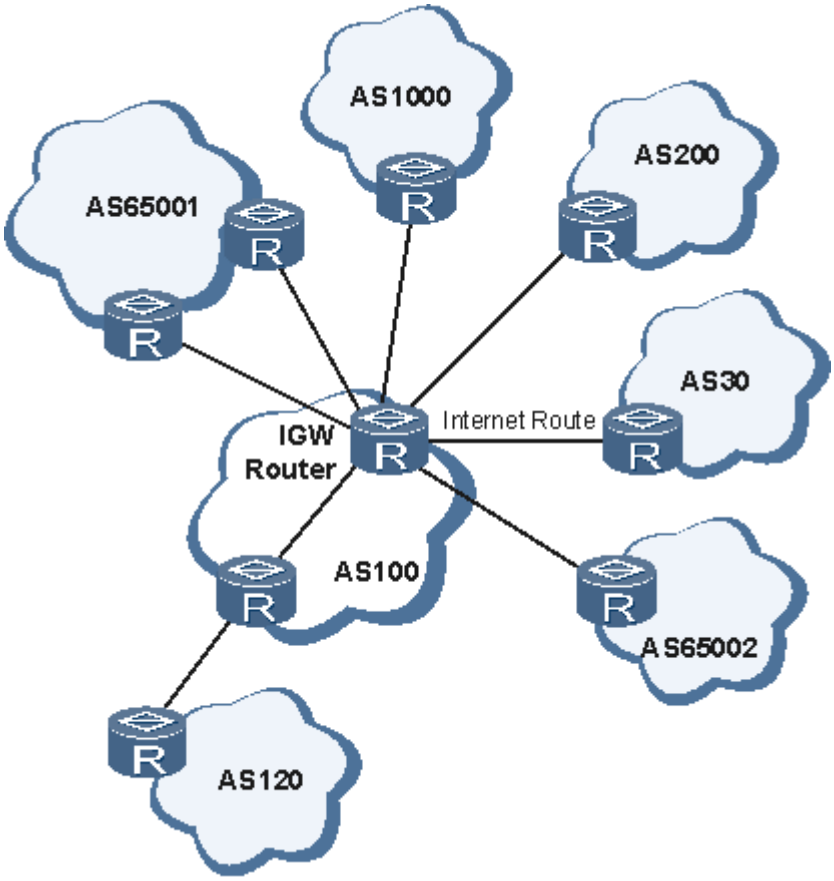
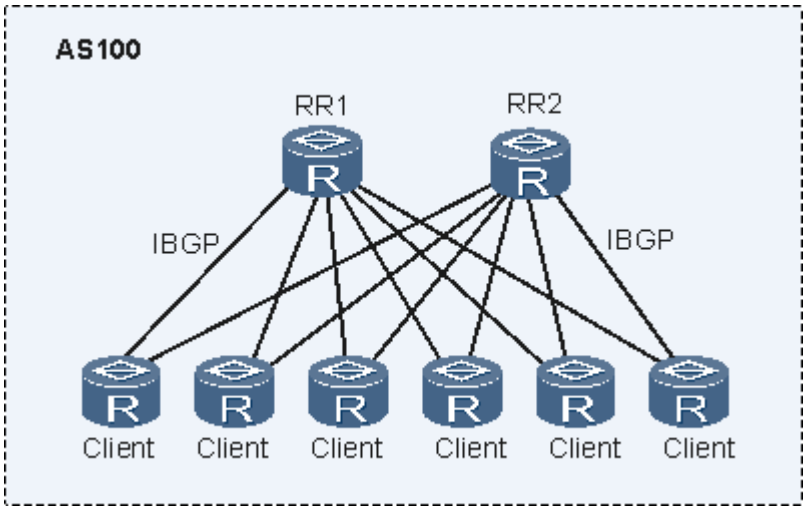
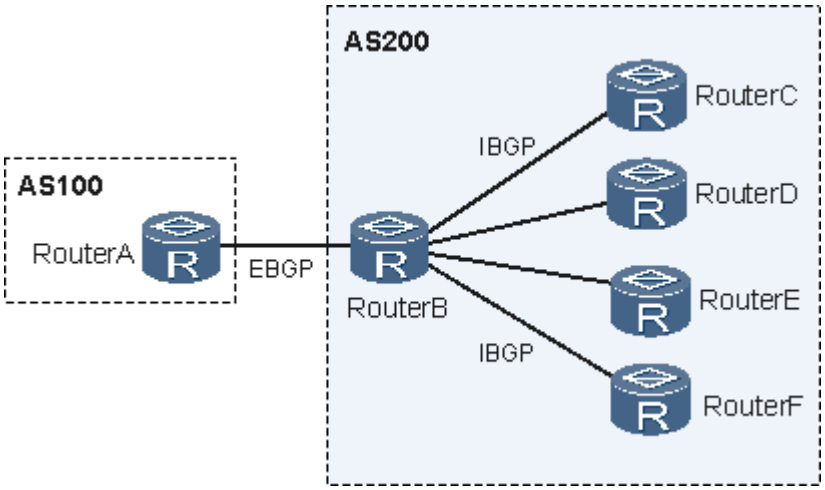


Figure 11-163 Networking for RRs with many clients



 Copyright © Huawei Technologies Co., Ltd.

Figure 11-164 Networking for a PE connected to multiple IBGP neighbors



The preceding scenarios have in common that a router needs to send routes to a large number of BGP peers, most of which share the same outbound policy. This situation is most evident in the networking shown in **Figure 11-163**.

For example, an RR has 100 clients and needs to reflect 100,000 routes to them. If the RR groups the routes for each peer before sending the routes to 100 clients, the total number of times that all routes are grouped is 100,000 x 100. After the dynamic update peer-groups feature is applied, the total number of times that all routes are grouped changes to 100,000 x 1. The efficiency is 100 times higher than before.

11.12.8.2.11 BGP NSR

The Border Gateway Protocol (BGP) non-stop routing (NSR) technique ensures the control plane connection to peers and uninterrupted traffic transmission on the forwarding plane if the BGP control plane fails because of active main board (AMB) failures.

Background

Carriers have increasing demands for IP network reliability. Conventional non-stop forwarding (NSF) and graceful restart (GR) techniques cannot prevent traffic interruptions if a peer does not support GR or multiple peers fail simultaneously during a GR process. Traffic is interrupted temporarily before the GR process is complete because a GR-enabled router cannot obtain routing information from or establish control plane connections to its peers during the GR process. To tackle this problem, NSR is introduced. NSR is an innovation, compared with NSF. NSR can be used to ensure uninterrupted traffic transmission and retain control plane connections if a software or hardware fault occurs on the control plane of a router. In addition, the fault is transparent to the control planes of its peers.

Related Concepts

- High availability (HA): supports data backup between the AMB and standby main board (SMB).

 Copyright © Huawei Technologies Co., Ltd.

- NSR: allows a standby control plane to take over traffic from an active control plane if the active control plane fails and prevents the control planes of peers from detecting the fault.
- NSF: enables a node to use the GR mechanism to ensure uninterrupted transmission during an AMB/SMB switchover.
- AMB and SMB: run control plane processes.

Implementation

BGP NSR implements synchronization of the following BGP data between the AMB and SMB to support NSR.

- BGP forwarding entries
- BGP control blocks

NOTE

The AMB and SMB must be installed on an NSR-enabled node to support NSR.

The NSR process is as follows:

1. Batch backup

   The AMB backs up BGP data in batches to the SMB immediately after the SMB starts.
2. Real-time backup

   The SMB backs up BGP data in real time and receives packets with the AMB simultaneously.
3. Switchover

   If the AMB fails, the SMB takes over services. The SMB retains uninterrupted operation of both the control and forwarding planes because data is synchronous between the AMB and SMB.

Other Usages

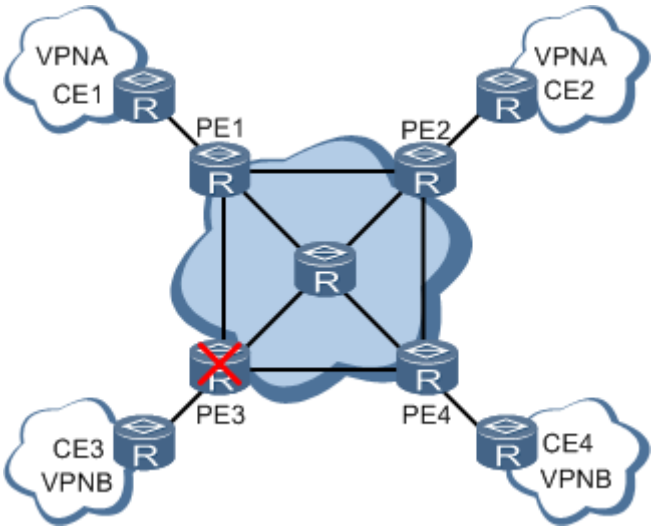
An NSR device can function as a GR Helper. The GR Helper communicates with NSR-disabled devices and responds to peers' GR requests during an AMB/SMB switchover.

Usage Scenario

- NSR is enabled on a node (for example, PE3 shown in **Figure 11-165**) with multiple links among which traffic is load-balanced. NSR helps to prevent traffic interruptions caused by a single point of failure. **Figure 11-165** shows a typical networking for NSR application.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-165 Typical networking for BGP NSR application



- NSR minimizes the impact of control plane faults and prevents route flapping on heavily loaded networks.

Benefits

- NSR ensures the uninterrupted operation of the forwarding plane and allows a standby control plane to retain connections between a local router and its peers if the local BGP control plane fails.
- With NSR, BGP routers can work independently without the assistance of peers. Even though control planes of multiple BGP routers fail, NSR can still ensure the AMB/SMB switchover on each router.

11.12.8.2.12 4-Byte AS Number

Purpose

2-byte autonomous system (AS) numbers used on networks range from 1 to 65535, and the available AS numbers are close to exhaustion as networks expand. Therefore, the AS number range needs to be extended. 4-byte AS numbers ranging from 1 to 4294967295 can address this problem. New speakers that support 4-byte AS numbers can co-exist with old speakers that support only 2-byte AS numbers.

Definition

4-byte AS numbers are extended from 2-byte AS numbers. Border Gateway Protocol (BGP) peers use a new capability code and optional transitive attributes to negotiate the 4-byte AS number capability and transmit 4-byte AS numbers. This mechanism enables communication between new speakers and between old speakers and new speakers.

Open capability code (0x41), defined by RFC 3392, indicates that the local end supports 4-byte capability extension.

The following new optional transitive attributes are defined by RFC 4893 and used to transmit 4-byte AS numbers in old sessions:

 Copyright © Huawei Technologies Co., Ltd.

- **AS4_Path** coded 0x11
- **AS4_Aggregator** coded 0x12

If a new speaker with an AS number greater than 65535 communicates with an old speaker, the old speaker needs to set the peer AS number to **AS_TRANS**. The value of **AS_TRANS** is 23456 and reserved.

Related Concepts

- New speaker: a peer that supports 4-byte AS numbers
- Old speaker: a peer that does not support 4-byte AS numbers
- New session: a BGP connection established between new speakers
- Old session: a BGP connection established between a new speaker and an old speaker, or between old speakers

Principles

BGP speakers negotiate capabilities by exchanging Open messages. **Figure 11-166** shows the format of Open messages exchanged between new speakers. The header of a BGP Open message is fixed, in which **My AS Number** is supposed to be the local AS number. However, **My AS Number** can carry only 2-byte AS numbers. Therefore, a new speaker adds 23456 to **My AS Number** and its local AS number to **Optional parameters** before it sends an Open message to a peer. After the peer receives the message, it can determine whether the new speaker supports 4-byte AS numbers by checking **Optional parameters** in the message.

Figure 11-166 Format of Open messages sent by new speakers

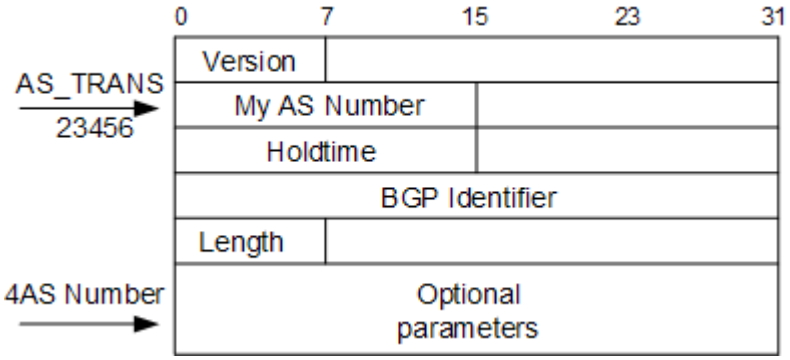
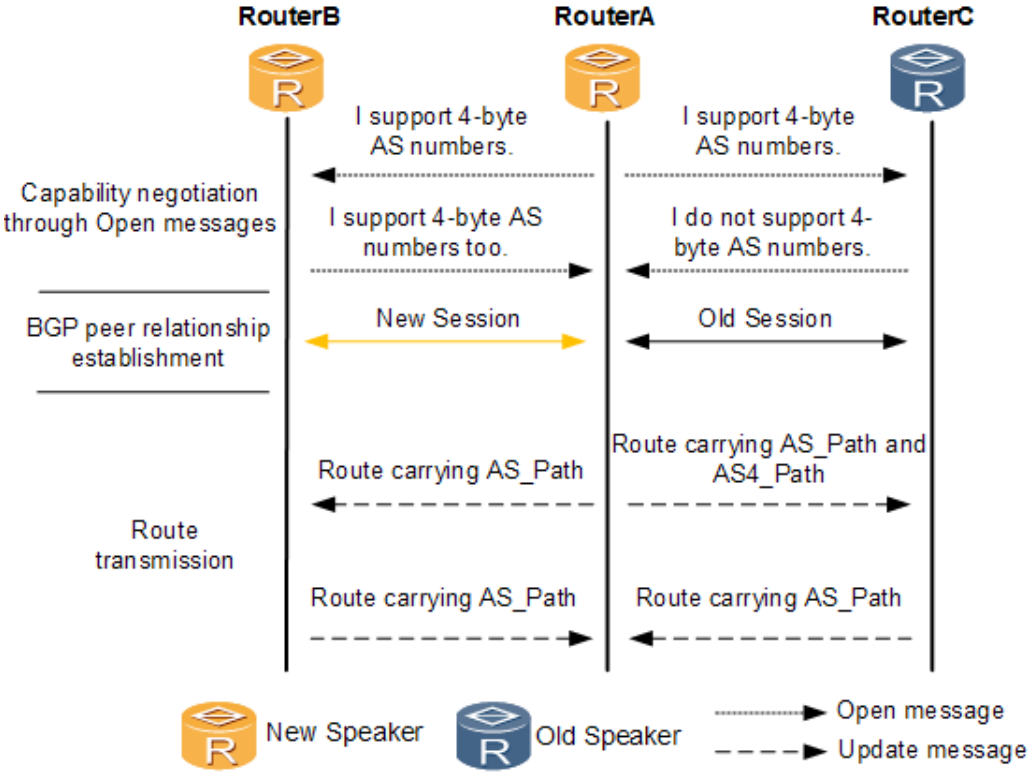


Figure 11-167 shows how peer relationships are established between new speakers, and between an old speaker and a new speaker. BGP speakers notify each other of whether they support 4-byte AS numbers by exchanging Open messages. After the capability negotiation, new sessions are established between new speakers, and old sessions are established between a new speaker and an old speaker.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-167 Process of establishing a BGP peer relationship

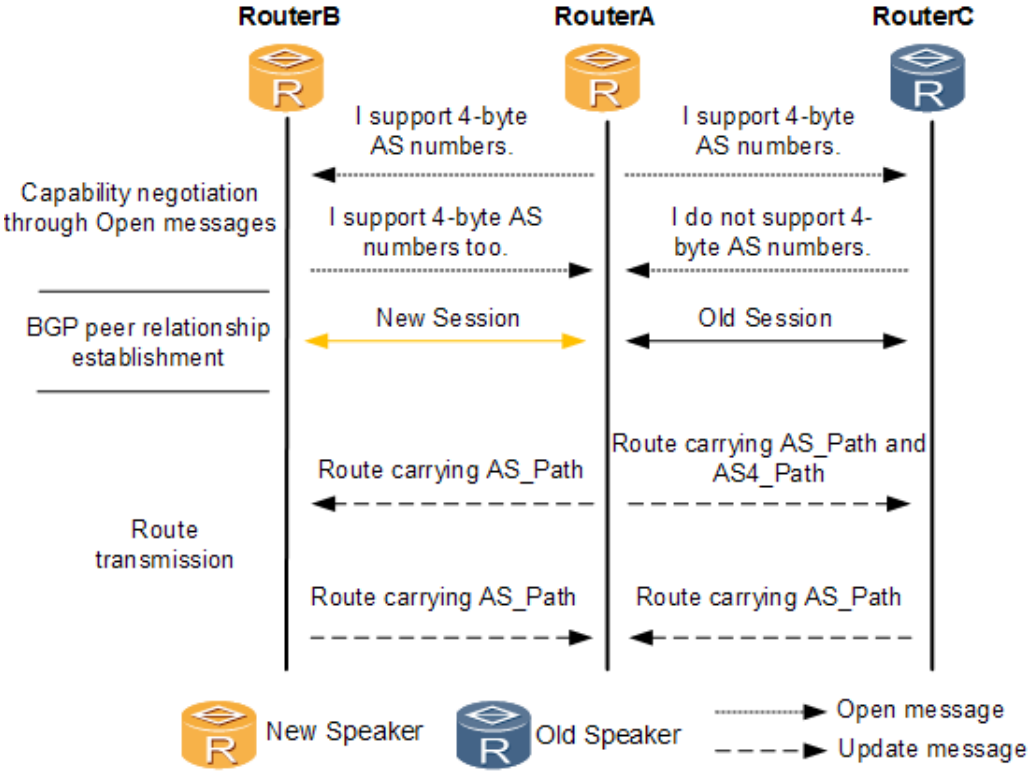


AS_Path and **Aggregator** in Update messages exchanged between new speakers carry 4-byte AS numbers, while **AS_Path** and **Aggregator** in Update messages sent by an old speaker carry 2-byte AS numbers.

- When a new speaker sends an Update message carrying an AS number greater than 65535 to an old speaker, the new speaker uses **AS4_Path** and **AS4_Aggregator** to assist **AS_Path** and **AS_Aggregator** in transferring 4-byte AS numbers. **AS4_Path** and **AS4_Aggregator** are transparent to the old speaker. In the networking shown in **Figure 11-168**, before the new speaker in AS 2.2 sends an Update message to the old speaker in AS 65002, the new speaker replaces each 4-byte AS number (1.1 and 2.2) with 23456 in **AS_Path**. Therefore, the **AS_Path** carried in the Update message is (23456, 23456, 65001), and the carried **AS4_Path** is (1.1, 2.2). After the old speaker in AS 65002 receives the Update message, it transparently transmits the message to other ASs.
- When the new speaker receives an Update message carrying **AS_Path**, **AS4_Path**, **AS_Aggregator**, and **AS4_Aggregator** from the old speaker, the new speaker uses the reconstruction algorithm to reconstruct the actual **AS_Path** and **AS_Aggregator**. In the networking shown in **Figure 11-168**, after the new speaker in AS 65003 receives an Update message carrying **AS_Path** (65002, 23456, 23456, 65001) and **AS4_Path** (1.1, 2.2) from the old speaker in AS 65002, the new speaker reconstructs the actual **AS_Path** (65002, 2.2, 1.1, 65001).

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-168 Process of transmitting a BGP Update message



Format of 4-byte AS numbers

A 4-byte AS number can be an integer or in dotted notation. The system stores 4-byte AS numbers as unsigned integers, regardless of their formats. 4-byte AS numbers in dotted notation are in the format of A.B. The formula of the conversion between 4-byte AS numbers for the two formats is as follows: Integer 4-byte AS number = A x 65536 + B. For example, the 4-byte AS number in dotted notation **2.3** can be converted to the integer 4-byte AS number 131075 (2 x 65536 + 3).

The supports 4-byte AS numbers of both formats. The 4-byte AS numbers displayed in the configuration files are in the format configured by users.

By default, the 4-byte AS numbers displayed in the **display** and **debugging** command outputs are in dotted notation, regardless of the configured format. If users change the default display format of 4-byte AS numbers from dotted notation to an integer, the displayed 4-byte AS numbers are integers.

NOTICE

Adjusting the display format of 4-byte AS numbers affects the matching results of **AS_Path** regular expressions and **extcommunity** filters. If you adjust the display format of 4-byte AS numbers on a system that uses an **AS_Path** regular expression or **extcommunity** filter as the export or import policy, reconfigure the **AS_Path** regular expression or **extcommunity** filter. If you do not reconfigure the **AS_Path** regular expression or **extcommunity** filter, routes cannot match the export or import policy, and a network error may occur.

 Copyright © Huawei Technologies Co., Ltd.

Benefits

4-byte AS numbers alleviate AS number exhaustion and therefore are beneficial to carriers who need to expand the network scale.

11.12.8.2.13 BGP Best-external

Background

If multiple routes to the same destination are available, a Border Gateway Protocol (BGP) device selects one optimal route based on BGP route selection policies and advertises the route to its BGP peers.

However, in scenarios with master and backup provider edges (PEs) or route reflectors (RRs), if routes are selected based on the preceding policies and the primary link fails, the BGP route convergence takes a long time because no backup route is available. To address this problem, the BGP Best-external feature was introduced.

Related Concepts

BGP Best-external: A mechanism that enables a backup device to select a sub-optimal route and send the route to its BGP peers if the route preferentially selected based on BGP route selection policies is an Internal Border Gateway Protocol (IBGP) route advertised by the master device. Therefore, BGP Best-external speeds up BGP route convergence if the primary link fails.

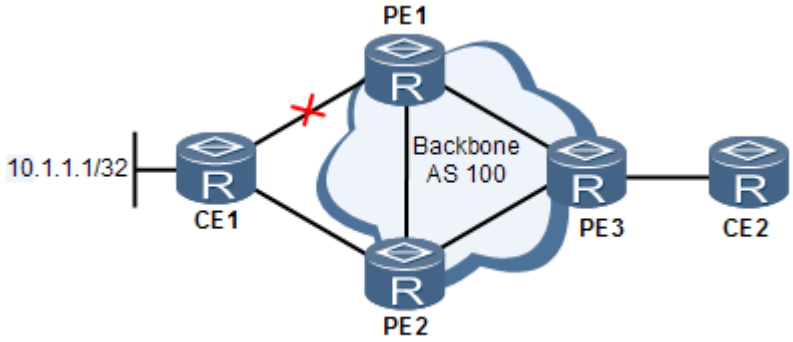
Best-external route: The sub-optimal route selected after BGP Best-external is enabled.

Networking with Master and Backup PEs

In the networking shown in **Figure 11-169**, CE1 is dual-homed to PE1 and PE2. PE1 has a greater Local_Pref value than PE2, and therefore PE1 is the master device while PE2 is the backup device. EBGP peer relationships are established between CE1 and PE1, and between CE1 and PE1. In addition, IBGP peer relationships are established among PE1, PE2, and PE3. PE1 and PE2 receive the same route to 10.1.1.1/32 from CE1. After receiving this route, PE1 advertises it to PE2 and PE3. Therefore, PE2 has two routes to 10.1.1.1/32. Of the two routes, PE2 preferentially selects the route from PE1 because PE1 has a larger Local_Pref value. PE2 does not advertise the selected route to PE3. Therefore, PE3 has only one route to 10.1.1.1/32, which is advertised by PE1. If the link between CE1 and PE1 fails, a new route must be selected to take over traffic after routes are converged, which takes a long time.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-169 Networking with master and backup PEs



BGP Best-external can be enabled on PE2 to address this problem. With BGP Best-external, PE2 selects the EBGP route from CE1 and advertises it to PE3. Therefore, PE3 has two routes to 10.1.1.1/32, in which the route CE1 -> PE2 -> PE3 backs up CE1 -> PE1 -> PE3. **Table 11-80** lists the differences with and without BGP Best-external.

Table 11-80 Differences with and without BGP Best-external

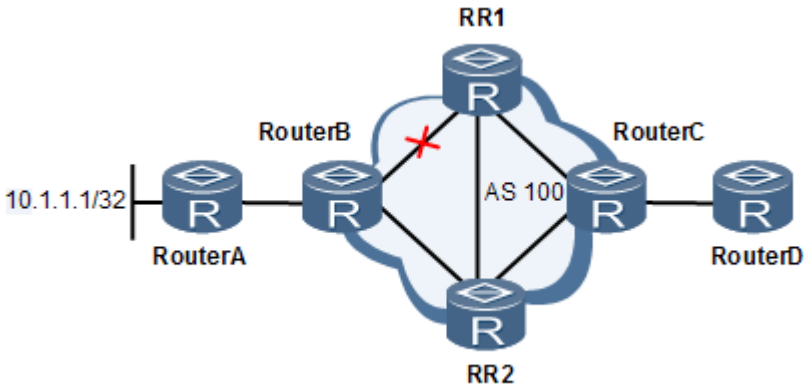
BGP Best-external	Route Available on PE3	Optimal Route	If the Link Between CE1 and PE1 Fails
Not enabled	CE1 -> PE1 -> PE3	CE1 -> PE1 -> PE3	A new route must be selected to take over traffic after routes are converged.
Enabled	CE1 -> PE1 -> PE3 CE1 -> PE2 -> PE3	CE1 -> PE1 -> PE3	Traffic is switched to CE1 -> PE2 -> PE3 immediately.

Networking with Master and Backup RRs

In the networking shown in **Figure 11-170**, an EBGP peer relationship is established between Router A and Router B. An IBGP peer relationship is established between every two devices among RR1, RR2, Router B, and Router C except between Router B and Router C. Router B is a client of RR1 and RR2. RR1 has a greater Local_Pref value than RR2, and therefore RR1 is the master device while RR2 is the backup device. RR1 and RR2 receive the same route to 10.1.1.1/32 from Router B. After receiving this route, RR1 advertises it to RR2 and Router C. Therefore, RR2 has two routes to 10.1.1.1/32. Of the two routes, RR2 preferentially selects the route from RR1 because RR1 has a greater Local_Pref value. RR2 does not advertise the selected route to Router C. Therefore, Router C has only one route to 10.1.1.1/32, which is advertised by RR1. If the link between Router B and RR1 fails, a new route must be selected to take over traffic after routes are converged, which takes a long time.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-170 Networking with master and backup RRs



BGP Best-external can be enabled on RR2 to address this problem. With BGP Best-external, RR2 selects the EBGP route from Router B and advertises it to Router C. Therefore, Router C has two routes to 10.1.1.1/32, in which the route Router A -> Router B -> RR2 -> Router C backs up Router A -> Router B -> RR1 -> Router C. **Table 11-81** lists the differences with and without BGP Best-external.

Table 11-81 Differences with and without BGP Best-external

BGP Best-external	Route Available on Router C	Optimal Route	If the Link Between Router B and RR1 Fails
Not enabled	Router A -> Router B -> RR1 -> Router C	Router A -> Router B -> RR1 -> Router C	A new route must be selected to take over traffic after routes are converged.
Enabled	Router A -> Router B -> RR1 -> Router C Router A -> Router B -> RR2 -> Router C	Router A -> Router B -> RR1 -> Router C	Traffic is switched to Router A -> Router B -> RR2 -> Router C immediately.

Usage Scenario

The BGP Best-external feature applies to scenarios in which master and backup PEs or RRs are deployed and the backup PE or RR needs to advertise the sub-optimal route (Best-external route) to its BGP peers to speed up BGP route convergence.

Advantages

As networks develop, services, such as voice over IP (VoIP), online video, and financial services, pose higher requirements for real-time transmission. With BGP Best-external, the backup device selects the sub-optimal route and advertises the

 Copyright © Huawei Technologies Co., Ltd.

route to its BGP peers, which speeds up BGP route convergence and minimizes service interruptions.

11.12.8.2.14 BGP ADD-PATH

Background

In a scenario with a route reflector (RR) and clients, if the RR has multiple routes to the same destination (with the same prefix), the RR selects an optimal route from these routes and then sends only the optimal route to its clients. Therefore, the clients have only one route to the destination. If a link along this route fails, route convergence takes a long time, which cannot meet the requirements on high reliability.

To address this issue, deploy the BGP ADD-PATH feature on the RR. With BGP ADD-PATH, the RR can send two or more routes with the same prefix to its clients. After reaching the clients, these routes can back up each other or load-balance traffic, which ensures high reliability in data transmission.

NOTE

- BGP ADD-PATH is deployed on RRs in most cases although it can be configured on any router.
- With BGP ADD-PATH, you can configure the maximum number of routes with the same prefix that an RR can send to its clients. The actual number of routes with the same prefix that an RR can send to its clients is the smaller value between the configured maximum number and the number of available routes with the same prefix.

Related Concepts

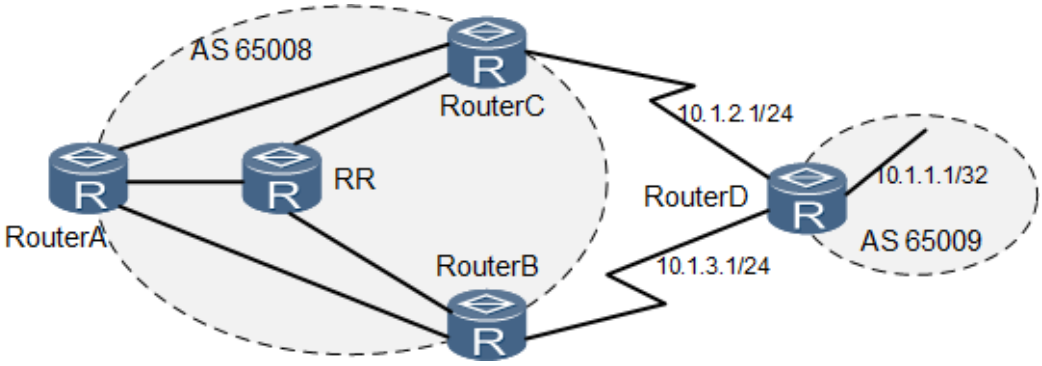
ADD-PATH route: The routes selected by BGP after BGP ADD-PATH is configured.

Typical Networking

On the network shown in **Figure 11-171**, Router A, Router B, and Router C are clients of the RR, and Router D is an EBGP peer of Router B and Router C.

Each of Router B and Router C receives a route to 10.1.1.1/32 from Router D, with 10.1.2.1/24 and 10.1.3.1/24 as the next hops, respectively. Then, each of Router B and Router C sends the received route to the RR. After receiving the two routes, the RR selects an optimal route from them and sends it to Router A. Therefore, Router A has only one route to 10.1.1.1/32.

Figure 11-171 Networking with BGP ADD-PATH



 Copyright © Huawei Technologies Co., Ltd.

BGP ADD-PATH can be configured to allow the RR to send more than one route with the same prefix to Router A. Suppose that the configured maximum number of routes with the same prefix that the RR can send to Router A is 2 and that the optimal route selected by the RR is the one from Router B. **Table 11-82** lists the differences with and without BGP ADD-PATH.

Table 11-82 Differences with and without BGP ADD-PATH

BGP ADD-PATH	Route Available on Router A	If the Link Between Router D and Router B Fails
Not enabled	Only one route to 10.1.1.1/32, with 10.1.2.1/24 as the next hop	A new route must be selected to take over traffic after route convergence.
Enabled	Two routes to 10.1.1.1/32, with 10.1.2.1/24 and 10.1.3.1/24 as the next hops	If the two routes back up each other, the link between Router D and Router C takes over traffic. If the two routes load-balance traffic, the link between Router D and Router C takes over all the traffic that ran along the two routes.

Usage Scenario

The BGP ADD-PATH feature applies to scenarios in which an RR and clients are deployed and the RR needs to send more than one route with the same prefix to its clients to ensure high reliability in data transmission.

Benefits

Deploying BGP ADD-PATH can improve network reliability.

11.12.8.2.15 IP FRR

Definition

IP fast reroute (IP FRR) provides fast traffic recovery upon link-layer or physical-layer faults. When a fault is detected and reported to the upper-layer routing system, a backup link is used to forward the impacted traffic.

IP FRR recovers the traffic impacted by link faults.

 Copyright © Huawei Technologies Co., Ltd.

Purpose

If a forwarding link fails on a traditional IP network, the device port goes down. After detecting such a fault, the device notifies the upper-layer routing system so that it updates the link status and recalculates a route. The process takes several seconds.

The convergence time cannot meet the requirements of delay-sensitive services, such as voice over IP (VoIP) and online video services. In contrast, IP FRR can recover impacted traffic in the level of 50 ms.

Principle

IP FRR supports static routes and dynamic routes, such as OSPF, OSPFv3, IS-IS, IS-IS IPv6, BGP, and BGP4+ routes.

IP FRR is implemented as follows:

1. Run commands to configure and enable IP FRR.

2. After IP FRR is enabled, the MA5800-X17 selects primary routes using routing protocols, calculates the backup links in advance, and saves them in the forwarding table.

3. If a primary link is unavailable, the MA5800-X17 forwards traffic through the backup link.

NOTE

For detailed configuration of IP FRR for each routing protocol, see the corresponding protocol description.

IP FRR Triggering Mode

IP FRR can be triggered by a BFD session or port status.

IP FRR triggered by a BFD session

If a Bidirectional Forwarding Detection (BFD) session is configured, it automatically associates with IP FRR. If a BFD session detects a link failure, the BFD session goes down and triggers IP FRR.

If multiple BFD sessions are configured and they have identical outgoing interfaces and destination IP addresses, IP FRR can be triggered if any of the BFD sessions goes down.

IP FRR triggered by port status

NOTE

This function is supported in V100R019C1X and later versions.

You can run the **port frr-binding** command to enable the function of triggering IP FRR by port status. When a port goes down, IP FRR is triggered.

The following restrictions apply when enabling the function of triggering IP FRR by port status:

- Virtual service ports do not support IP FRR. Only standard VLAN ports support IP FRR.

 Copyright © Huawei Technologies Co., Ltd.

- To enable the function of triggering IP FRR by port status in a VLAN, the VLAN can have only one standard port, which can be a common Ethernet port or a port aggregation group.
- Protection groups do not support IP FRR.

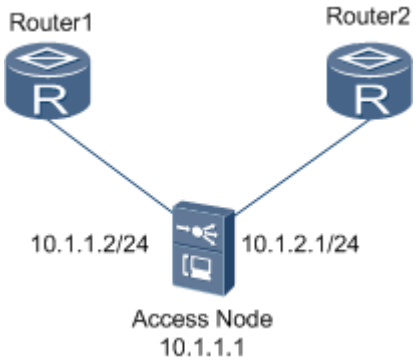
11.12.8.3 Configuration Example of BGP

This topic provides an example for configuring the BGP on the device.

Service Requirements

In this example network, an IBGP connection is set up between access node and routers.

Figure 11-172 Example network for configuring the BGP



Data Plan

Table 11-83 provides the data plan for configuring the BGP.

Table 11-83 Data plan for configuring the BGP

Item	Data	Remarks
Access node	IP address of VLAN interface 6: 10.1.1.2/24	It is used for the IBGP connection to the Router1.
	IP address of VLAN interface 2: 10.1.2.1/24	It is used for the IBGP connection to the Router12.
	Router ID: 10.1.1.1	-
	AS number: 2000	-

Procedure

Step 1 Configure access node.

 Copyright © Huawei Technologies Co., Ltd.

1. Configure the IP address of the Layer 3 interface.

```
huawei(config)#vlan 6 smart
huawei(config)#port vlan 6 0/9 0
huawei(config)#interface vlanif 6
huawei(config-if-vlanif6)#ip address 10.1.1.2 24
huawei(config-if-vlanif6)#quit
huawei(config)#vlan 2 smart
huawei(config)#port vlan 2 0/9 1
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.1.2.1 24
huawei(config-if-vlanif2)#quit
```

2. Enable the BGP function.

```
huawei(config)#bgp 2000
huawei(config-bgp)#router-id 10.1.1.1
huawei(config-bgp)#peer 10.1.3.3 as-number 2000
huawei(config-bgp)#peer 10.1.2.2 as-number 2000
huawei(config-bgp)#quit
```

3. Save the data.

```
huawei(config)#save
```

----End

Result

- Run the **display bgp peer** command, and you can query the configuration.

Configuration File

```
vlan 6 smart
port vlan 6 0/9 0
interface vlanif 6
ip address 10.1.1.2 24
quit
vlan 2 smart
port vlan 2 0/9 0
interface vlanif 2
ip address 10.1.2.1 24
quit
bgp 2000
router-id 10.1.1.1
peer 10.1.3.3 as-number 2000
peer 10.1.2.2 as-number 2000
quit
save
```

11.12.8.4 Configuration Example of BGP4+

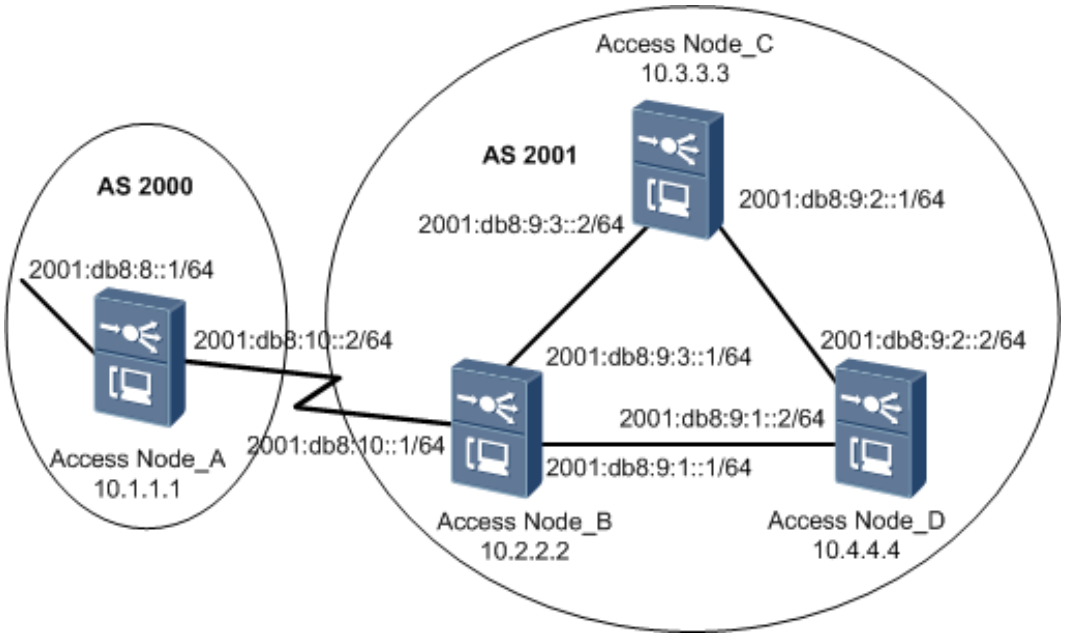
This topic provides an example for configuring the BGP4+ on the device.

Service Requirements

In this example network, an external Border Gateway Protocol (EBGP) connection is set up between Access node_A and Access node_B, and an Interior Border Gateway Protocol (IBGP) connection is set up among Access node_B, Access node_C, and Access node_D.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-173 Example network for configuring the BGP4+



Data Plan

Table 11-84 provides the data plan for configuring the BGP4+.

Table 11-84 Data plan for configuring the BGP4+

Item	Data	Remarks
Access node_A	IPv6 address of virtual local area network (VLAN) interface 6: 2001:db8:10::2/64	It is used for the EBGP connection to Autonomous System (AS) 2001.
	IPv6 address of VLAN interface 2: 2001:db8:8::1/64	-
	Router ID: 10.1.1.1	-
	AS number: 2000	-
Access node_B	IPv6 address of VLAN interface 6: 2001:db8:10::1/64	It is used for the EBGP connection to AS 2000.
	IPv6 address of VLAN interface 3: 2001:db8:9:3::1/64	It is used for the IBGP connection to the Access node_C.
	IPv6 address of VLAN interface 4: 2001:db8:9:1::1/64	It is used for the IBGP connection to the Access node_D.
	Router ID: 10.2.2.2	-

 Copyright © Huawei Technologies Co., Ltd.

Item	Data	Remarks
	AS number: 2001	-
Access node_C	IPv6 address of VLAN interface 3: 2001:db8:9:3::2/64	It is used for the IBGP connection to the Access node_B.
	IPv6 address of VLAN interface 4: 2001:db8:9:2::1/64	It is used for the IBGP connection to the Access node_D.
	Router ID: 10.3.3.3	-
	AS number: 2001	-
Access node_D	IPv6 address of VLAN interface 5: 2001:db8:9:2::2/64	It is used for the IBGP connection to the Access node_C.
	IPv6 address of VLAN interface 4: 2001:db8:9:1::2/64	It is used for the IBGP connection to the Access node_B.
	Router ID: 10.4.4.4	-
	AS number: 2001	-

Procedure

Step 1 Configure Access node_A.

1. Configure the IPv6 address of the Layer 3 interface.

```
huawei(config)#ipv6
huawei(config)#vlan 6 smart
huawei(config)#port vlan 6 0/9 0
huawei(config)#interface vlanif 6
huawei(config-if-vlanif6)#ipv6 enable
huawei(config-if-vlanif6)#ipv6 address 2001:db8:10::2 64
huawei(config-if-vlanif6)#quit
huawei(config)#vlan 2 smart
huawei(config)#port vlan 2 0/9 1
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ipv6 enable
huawei(config-if-vlanif2)#ipv6 address 2001:db8:8::1/64
huawei(config-if-vlanif2)#quit
```

2. Enable the Border Gateway Protocol (BGP) function and configure the EBGP neighbor between Access node_B.

```
huawei(config)#bgp 2000
huawei(config-bgp)#router-id 10.1.1.1
huawei(config-bgp)#peer 2001:db8:10::1 as-number 2001
huawei(config-bgp)#ipv6-family unicast
huawei(config-bgp-af-ipv6)#peer 2001:db8:10::1 enable
huawei(config-bgp-af-ipv6)#network 2001:db8:10:: 64
huawei(config-bgp-af-ipv6)#network 2001:db8:8:: 64
huawei(config-bgp-af-ipv6)#quit
huawei(config-bgp)#quit
```

3. Save the data.

```
huawei(config)#save
```

 Copyright © Huawei Technologies Co., Ltd.

Step 2 Configure Access node_B.

1. Configure the IPv6 address of the Layer 3 interface.

```
huawei(config)#ipv6
huawei(config)#vlan 6 smart
huawei(config)#port vlan 6 0/9 0
huawei(config)#interface vlanif 6
huawei(config-if-vlanif6)#ipv6 enable
huawei(config-if-vlanif6)#ipv6 address 2001:db8:10::1 64
huawei(config-if-vlanif6)#quit
huawei(config)#vlan 3 smart
huawei(config)#port vlan 3 0/9 0
huawei(config)#interface vlanif 3
huawei(config-if-vlanif3)#ipv6 enable
huawei(config-if-vlanif3)#ipv6 address 2001:db8:9:3::1 64
huawei(config-if-vlanif3)#quit
huawei(config)#vlan 4 smart
huawei(config)#port vlan 4 0/9 0
huawei(config)#interface vlanif 4
huawei(config-if-vlanif4)#ipv6 enable
huawei(config-if-vlanif4)#ipv6 address 2001:db8:9:1::1 64
huawei(config-if-vlanif4)#quit
```

2. Enable the BGP function. Configure the EBGP neighbor between Access node_B and Access node_A, and the IBGP neighbor between Access node_B, Access node_C, and Access node_D.

```
huawei(config)#bgp 2001
huawei(config-bgp)#router-id 10.2.2.2
huawei(config-bgp)#peer 2001:db8:10::2 as-number 2000
huawei(config-bgp)#peer 2001:db8:9:3::2 as-number 2001
huawei(config-bgp)#peer 2001:db8:9:1::2 as-number 2001
huawei(config-bgp)#ipv6-family unicast
huawei(config-bgp-af-ipv6)#peer 2001:db8:10::2 enable
huawei(config-bgp-af-ipv6)#peer 2001:db8:9:3::2 enable
huawei(config-bgp-af-ipv6)#peer 2001:db8:9:1::2 enable
huawei(config-bgp-af-ipv6)#import-route direct
huawei(config-bgp-af-ipv6)#quit
huawei(config-bgp)#quit
```

3. Save the data.

```
huawei(config)#save
```

Step 3 Configure Access node_C.

1. Configure the IPv6 address of the Layer 3 interface.

```
huawei(config)#ipv6
huawei(config)#vlan 3 smart
huawei(config)#port vlan 3 0/9 0
huawei(config)#interface vlanif 3
huawei(config-if-vlanif3)#ipv6 enable
huawei(config-if-vlanif3)#ipv6 address 2001:db8:9:3::2 64
huawei(config-if-vlanif3)#quit
huawei(config)#vlan 5 smart
huawei(config)#port vlan 5 0/9 0
huawei(config)#interface vlanif 5
huawei(config-if-vlanif5)#ipv6 enable
huawei(config-if-vlanif5)#ipv6 address 2001:db8:9:2::1 64
huawei(config-if-vlanif5)#quit
```

2. Enable the BGP function. Configure the IBGP neighbor between Access node_B and Access node_D.

```
huawei(config)#bgp 2001
huawei(config-bgp)#router-id 10.3.3.3
huawei(config-bgp)#peer 2001:db8:9:3::1 as-number 2001
huawei(config-bgp)#peer 2001:db8:9:2::2 as-number 2001
huawei(config-bgp)#ipv6-family unicast
huawei(config-bgp-af-ipv6)#peer 2001:db8:9:3::1 enable
huawei(config-bgp-af-ipv6)#peer 2001:db8:9:2::2 enable
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-bgp-af-ipv6)#import-route direct
huawei(config-bgp-af-ipv6)#quit
huawei(config-bgp)#quit
```

3. Save the data.

```
huawei(config)#save
```

Step 4 Configure Access node_D.

1. Configure the IPv6 address of the Layer 3 interface.

```
huawei(config)#ipv6
huawei(config)#vlan 4 smart
huawei(config)#port vlan 4 0/9 0
huawei(config)#interface vlanif 4
huawei(config-if-vlanif4)#ipv6 enable
huawei(config-if-vlanif4)#ipv6 address 2001:db8:9:1::2 64
huawei(config-if-vlanif4)#quit
huawei(config)#vlan 5 smart
huawei(config)#port vlan 5 0/9 0
huawei(config)#interface vlanif 5
huawei(config-if-vlanif5)#ipv6 enable
huawei(config-if-vlanif5)#ipv6 address 2001:db8:9:2::2 64
huawei(config-if-vlanif5)#quit
```

2. Enable the BGP function. Configure the IBGP neighbor between Access node_B and Access node_C.

```
huawei(config)#bgp 2001
huawei(config-bgp)#router-id 10.4.4.4
huawei(config-bgp)#peer 2001:db8:9:1::2 as-number 2001
huawei(config-bgp)#peer 2001:db8:9:2::1 as-number 2001
huawei(config-bgp)#ipv6-family unicast
huawei(config-bgp-af-ipv6)#peer 2001:db8:9:1::2 enable
huawei(config-bgp-af-ipv6)#peer 2001:db8:9:2::1 enable
huawei(config-bgp-af-ipv6)#import-route direct
huawei(config-bgp-af-ipv6)#quit
huawei(config-bgp)#quit
```

3. Save the data.

```
huawei(config)#save
```

----End

Result

- Run the **display bgp peer** command, and you can see that:
  - The EBGP connection is set up between Access node_A and Access node_B.
  - The IBGP connections are set up among Access node_B, Access node_C, and Access node_D.
  - The route with the destination subnet 2001:db8:8::/64 exists on Access node_C and Access node_D, and the next hop of the route is the interface address of Access node_A
- Run the **ping ipv6** command on Access node_C and Access node_D to ping the Layer 3 interface (2001:db8:8::1/64) on Access node_A. The **ping ipv6** command is executed successfully.

Configuration File

Configuration on each Access node is similar. Take Access Node_A for example.

```
ipv6
vlan 6 smart
```

 Copyright © Huawei Technologies Co., Ltd.

```
port vlan 6 0/9 0
interface vlanif 6
ipv6 enable
ipv6 address 2001:db8:10::2 64
quit
vlan 2 smart
port vlan 2 0/9 0
interface vlanif 2
ipv6 enable
ipv6 address 2001:db8:8::1 64
quit
bgp 2000
router-id 10.1.1.1
peer 2001:db8:10::1 as-number 2001
ipv6-family unicast
peer 2001:db8:10::1 enable
network 2001:db8:10:: 64
network 2001:db8:8:: 64
quit
quit
```

11.12.8.5 BGP Reference Standards and Protocols

The following table lists the references.

Table 11-85 References

Document No.	Document Name
RFC 827	Exterior Gateway Protocol (EGP)
RFC 1213	Management Information Base for Network Management of TCP/IP-based internets: MIB-II
RFC 1397	Default Route Advertisement in BGP2 and BGP3 Version of the Border Gateway Protocol
RFC 1403	BGP OSPF Interaction
RFC 1519	Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy
RFC 1654	A Border Gateway Protocol 4 (BGP-4)
RFC 1655	Application of the Border Gateway Protocol in the Internet
RFC 1656	BGP-4 Protocol Document Roadmap and Implementation Experience
RFC 1657	basic BGP4 MIB
RFC 1771	A Border Gateway Protocol 4 (BGP-4)
RFC 1772	BGP basic functions support
RFC 1773	obsoletes RFC 1656
RFC 1774	BGP-4 Protocol Analysis
RFC 1930	Guidelines for creation, selection, and registration of an Autonomous System (AS)

 Copyright © Huawei Technologies Co., Ltd.

Document No.	Document Name
RFC 1965	Autonomous System Confederations for BGP
RFC 1966	BGP Route-Reflection
RFC 1997	BGP Communities Attribute
RFC 1998	An Application of the BGP Community Attribute
RFC 2270	Using a Dedicated AS for Sites Homed to a Single Provider
RFC 2283	Multiprotocol Extensions for BGP-4
RFC 2385	TCP MD5
RFC 2439	BGP Route Flap Damping
RFC 2519	A Framework for Inter-Domain Route Aggregation
RFC 2545	BGP supporting IPV6
RFC 2796	BGP Route Reflection
RFC 2842	Capabilities Advertisement with BGP-4
RFC 2858	Multiprotocol Extensions for BGP-4
RFC 2918	Route Refresh Capability for BGP-4
RFC 3065	Autonomous System Confederations for BGP
RFC 3107	Support BGP carry Label for MPLS
RFC 3232	Assigned Numbers: RFC 1700 is Replaced by an On-line Database
RFC 3392	Capabilities Advertisement with BGP-4
RFC 3562	Key Management Considerations for the TCP MD5 Signature Option
RFC 4271	A Border Gateway Protocol 4 (BGP-4)
RFC 4272	BGP Security Vulnerabilities Analysis
RFC 4273	Definitions of Managed Objects for the Fourth Version of Border Gateway
RFC 4274	BGP-4 Protocol Analysis
RFC 4276	BGP 4 Implementation Report
RFC 4277	Experience with the BGP-4 Protocol
RFC 4360	BGP Extended Communities Attribute
RFC 4456	BGP Route Reflection
RFC 4486	Subcodes for BGP Cease Notification Message

 Copyright © Huawei Technologies Co., Ltd.

Document No.	Document Name
RFC 4632	Classless Inter-domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan
RFC 4724	Graceful Restart Mechanism for BGP
RFC 4760	Multiprotocol Extensions for BGP-4
RFC 4781	Graceful Restart Mechanism for BGP with MPLS
RFC 4893	BGP Support for Four-octet AS Number Space
RFC 5004	Avoid BGP Best Path Transitions from One External to Another
RFC 5065	Autonomous System Confederations for BGP
RFC 5291	Outbound Route Filtering Capability for BGP-4
RFC 5292	Address-Prefix-Based Outbound Route Filter for BGP-4
RFC 5396	Textual Representation of Autonomous System (AS) Numbers
RFC 5398	Autonomous System (AS) Number Reservation for Documentation Use
RFC 5492	Capabilities Advertisement with BGP-4
draft-ietf-idr-as4bytes-13	BGP Support for Four-octet AS Number Space
draft-ietf-idr-bgp-ext-communities-05	Extended Community Attribute
draft-ietf-idr-as-pathlimit-03	The AS_PATHLIMIT Path Attribute
draft-ietf-idr-aspath-orf-09	Aspath Based Outbound Route Filter for BGP-4
Draft-marques-l3vpn-ibgp-01	Internal BGP as PE-CE protocol draft-marques-l3vpn-ibgp-01
draft-ramachandra-bgp-ext-communities-04	Address Prefix Based Outbound Route Filter for BGP-4
draft-kato-bgp-ipv6-link-local-00	BGP4+ Peering Using IPv6 Link-local Address
draft-ietf-idr-cap-neg-01	Capabilities Negotiation with BGP4
draft-ietf-idr-bgp-prefix-orf-04	Address Prefix Based Outbound Route Filter for BGP-4

 Copyright © Huawei Technologies Co., Ltd.

Document No.	Document Name
draft-ietf-idr-rfc3065bis-06	Autonomous System Confederations for BGP

11.12.9 Routing Policy

11.12.9.1 Introduction to Routing Policies

Definition

Routing policies are used to filter routes and set attributes for routes. Changing route attributes (including reachability) changes the path that network traffic passes through.

Purpose

When advertising, receiving, and importing routes, the Router implements certain policies based on actual networking requirements to filter routes and change the attributes of the routes. Routing policies serve the following purposes:

- Control route advertising

  Only routes that match the rules specified in a policy are advertised.
- Control route receiving

  Only the required and valid routes are received. This reduces the size of the routing table and improves network security.
- Filter and control imported routes

  A routing protocol may import routes discovered by other routing protocols. Only routes that satisfy certain conditions are imported to meet the requirements of the protocol.
- Modify attributes of specified routes

  Attributes of the routes that are filtered by a routing policy are modified to meet the requirements of the local device.

Benefits

This feature brings the following benefits:

- Controls the size of the routing table, saving system resources.
- Controls route receiving and advertising, improving network security.
- Modifies attributes of routes for proper traffic planning, improving network performance.

 Copyright © Huawei Technologies Co., Ltd.

11.12.9.2 Principles

Implementation

Routing policies are implemented using the following procedures:

- Define rules: Define features of routes to which routing policies are applied. Users define a set of matching rules based on different attributes of routes, such as the destination address and the address of the router that advertises the routes.
- Implement the rules: Apply the matching rules to routing policies for advertising, receiving, and importing routes.

Filter

Filters are the core of a routing policy and is defined using a set of matching rules. The provides several filters for routing policies.

Table 11-86 lists applicable scopes and matching rules of different filters.

Table 11-86 Comparisons between filters

Filter	Applicable Scope	Matching Rules
Access control list (ACL)	Dynamic routing protocols	Inbound interface, source or destination IP address, protocol type, and source or destination port number
IP prefix list	Dynamic routing protocols	Source and destination IP addresses and next hop address
Route-Policy	Dynamic routing protocols	Destination IP address, next hop address, cost, interface information, route type, ACL, IP prefix list, AS_Path filter, community filter, extcommunity filter, and RD filter.

The ACL and IP prefix list can be used only to filter routes but not modify attributes of the filtered routes. A Route-Policy is a comprehensive filter, and it can use the matching rules of the ACL and IP prefix list to filter routes. In addition, attributes of the filtered routes can be modified using the Route-Policy. The following section describes the filters in more detail.

ACL

An ACL is a set of sequential filtering rules. Users can define rules based on packet information, such as inbound interfaces, source or destination IP addresses, protocol types, or source or destination port numbers and specify an action to deny or permit packets. After an ACL is configured, the system classifies received packets based on the rules defined in the ACL and denies or permits the packets accordingly.

 Copyright © Huawei Technologies Co., Ltd.

An ACL only classifies packets based on defined rules and can be used to filter packets only when it is applied to a routing policy.

ACLs can be configured for both IPv4 packets and IPv6 packets. Based on the usage, ACLs are classified into three types: interface-based ACLs, basic ACLs, and advanced ACLs. Users can specify the IP address and subnet address range in an ACL to match the source IP address, destination network segment address, or the next hop address of a route.

ACLs can be configured on access or core devices to:

- Protect the devices against IP, TCP, and Internet Control Message Protocol (ICMP) packet attacks.
- Control network access. For example, ACLs can be used to control the access of enterprise network users to external networks, the specific network resources that users can access, and the period for which users can access networks.
- Limit network traffic and improve network performance. For example, ACLs can be used to limit bandwidth for upstream and downstream traffic, charge for the bandwidth that users have applied for, and fully use high-bandwidth network resources.

IP Prefix List

An IP prefix list contains a group of route filtering rules. Users can specify the prefix and mask length range to match the destination network segment address or the next hop address of a route. An IP prefix list is used to filter routes that are advertised and received by various dynamic routing protocols.

An IP prefix list is easier and more flexible than an ACL. However, if a large number of routes with different prefixes need to be filtered, configuring an IP prefix list to filter the routes is complex.

IP prefix lists can be configured for both IPv4 routes and IPv6 routes, and they share the same implementation process. An IP prefix list filters routes based on the mask length or mask length range.

- Mask length: An IP prefix list filters routes based on IP address prefixes. An IP address prefix is defined by an IP address and the mask length. For example, in a route to 10.1.1.1/16, the mask length is 16 bits, and the valid prefix is 16 bits (10.1.0.0).
- Mask length range: Routes with the IP address prefix and mask length within the range defined in the IP prefix list meet the matching rules.

NOTE

0.0.0.0 is a wildcard address. If the IP prefix is 0.0.0.0, users must specify either a mask or a mask length range, with the following results:

- If a mask is specified, all routes with the mask are permitted or denied as required.
- If a mask length range is specified, all routes with the mask length in the range are permitted or denied as required.

 Copyright © Huawei Technologies Co., Ltd.

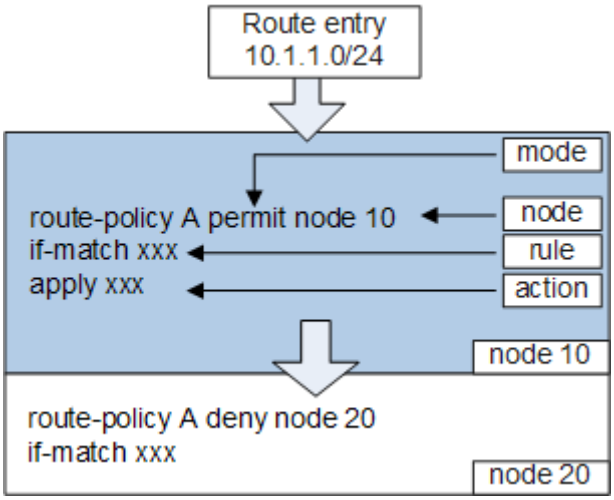
Route-Policy

A Route-Policy is a complex filter. It is used to match attributes of specified routes and change route attributes when specific conditions are met. A Route-Policy can use the preceding six filters to define its matching rules.

- **Composition of a Route-Policy**

  As shown in **Figure 11-174**, a Route-Policy consists of node IDs, matching mode, **if-match** clauses, and **apply** clauses.

  Figure 11-174 Composition of a Route-Policy



  - Node ID

    A Route-Policy consists of one or more nodes. Node IDs are specified as indexes in the IP prefix list. In a Route-Policy, routes are filtered based on the following rules:

    - Sequential matching: The system checks entries based on node IDs in ascending order. Therefore, specifying the node IDs in the required sequence is recommended.
    - One-time matching: The relationship between the nodes of a Route-Policy is "OR". If a route matches one node, the route matches the Route-Policy and will not be matched against the next node.

  - Matching mode

    Either of the following matching modes can be used:

    - **permit**: specifies the **permit** mode of a node. If a route matches the **if-match** clauses of a node, all the actions defined by **apply** clauses are performed, and the matching is complete. If a route does not match the **if-match** clauses of the node, the route continues to match the next node.
    - **deny**: specifies the **deny** mode of a node. In the **deny** mode, the **apply** clauses are not used. If a route matches all the **if-match** clauses of the node, the route is denied by the node and the next

 Copyright © Huawei Technologies Co., Ltd.

node is not matched. If the entry does not match all the **if-match** clauses, the next node is matched.

NOTE

To allow other routes to pass through, a Route-Policy that contains no **if-match** or **apply** clause in the **permit** mode needs to be configured for a node next to multiple nodes that are in the **deny** mode.

- **if-match** clause

  The **if-match** clause defines the matching rules.

  Each node of a Route-Policy can comprise multiple **if-match** clauses or no **if-match** clause at all. If no **if-match** clause is configured for a node in the **permit** mode, all routes match the node.

- **apply** clause

  The **apply** clauses specify actions. When a route matches a Route-Policy, the system sets some attributes for the route based on the **apply** clause.

  Each node of a Route-Policy can comprise multiple **apply** clauses or no **apply** clause at all. The **apply** clause is not used when routes need to be filtered but attributes of the routes do not need to be set.

- **Matching results of a Route-Policy**

  The matching results of a Route-Policy are obtained based on the following aspects:

  - Matching mode of the node, either **permit** or **deny**
  - Matching rules (either **permit** or **deny**) contained in the **if-match** clause (such as ACLs or IP prefix lists)

  The matching results are listed in **Table 11-87**.

Table 11-87 Matching results of a Route-Policy

Rule (Matching Rule Contained in if-match Clauses)	Mode (Matching Mode of a Node)	Matching Result
permit	**permit**	• Routes matching the **if-match** clauses of the node match the Route-Policy, and the matching is complete. • Routes not matching the **if-match** clauses of the node continue to match the next node of the Route-Policy.

 Copyright © Huawei Technologies Co., Ltd.

Rule (Matching Rule Contained in if-match Clauses)	Mode (Matching Mode of a Node)	Matching Result
	deny	• Routes matching the **if-match** clauses of the node are denied by the Route-Policy, and the matching is complete. • Routes not matching the **if-match** clauses of the node continue to match the next node of the Route-Policy.
deny	**permit**	• Routes matching the **if-match** clauses of the node are denied by the Route-Policy and continue to match the next node. • Routes not matching the **if-match** clauses of the node continue to match the next node of the Route-Policy.
	deny	• Routes matching the **if-match** clauses of the node are denied by the Route-Policy and continue to match the next node. • Routes not matching the **if-match** clauses of the node continue to match the next node of the Route-Policy. **NOTE** If all **if-match** clauses and nodes of the Route-Policy are in the **deny** mode, all the routes to be filtered are denied by the Route-Policy.

NOTE

All unmatched routes are denied by the Route-Policy by default. If more than one node is defined in a Route-Policy, at least one of them must be in the **permit** mode. The reason is as follows:

- If a route fails to match any of the nodes, the route is denied by the Route-Policy.
- If all the nodes in the Route-Policy are set in the **deny** mode, all the routes to be filtered are denied by the Route-Policy.

 Copyright © Huawei Technologies Co., Ltd.

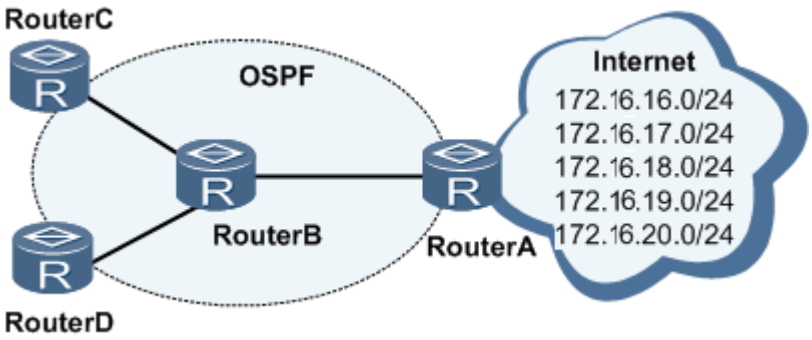
11.12.9.3 Applications

Specific Routes Filtering

On the OSPF-enabled network shown in **Figure 11-175**, Router A receives routes from the Internet and advertises some of the routes to Router B.

- Router A advertises only routes 172.16.17.0/24, 172.16.18.0/24, and 172.16.19.0/24 to Router B.
- Router C accepts only the route 172.16.18.0/24.
- Router D accepts all the routes advertised by Router B.

Figure 11-175 Networking diagram for filtering received and advertised routes



There are multiple approaches to meet the preceding requirements, and the following two approaches are used in this example:

- Use IP prefix lists.
  - Configure an IP prefix list for Router A and configure the IP prefix list as an export policy of Router A to be used by OSPF.
  - Configure another IP prefix list for Router C and configure the IP prefix list as an import policy of Router C to be used by OSPF.
- Use route-policies
  - Configure a Route-Policy (the matching rules can be the IP prefix list, cost, or route tag) for Router A and configure the Route-Policy as an export policy of Router A to be used by OSPF.
  - Configure another Route-Policy for Router C and configure the Route-Policy as an import policy of Router C to be used by OSPF.

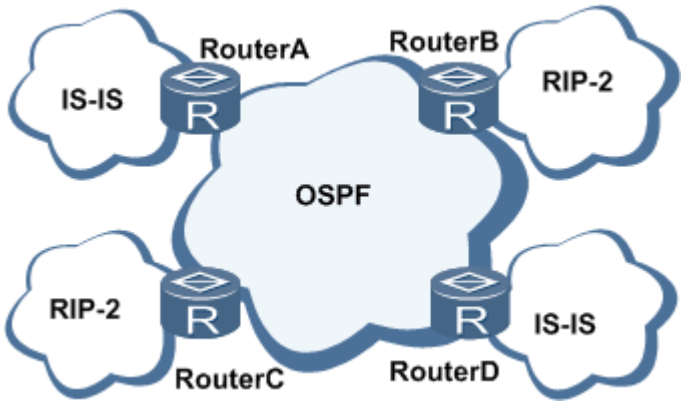
  Compared with an IP prefix list, a Route-Policy allows route attributes to be modified and can be used to control routes more flexibly, but it is more complex to configure.

Transparent Transmission of Routes of Other Protocols Through an OSPF AS

On the network shown in **Figure 11-176**, an AS runs OSPF and functions as a transit AS for other areas. Routes from the IS-IS area connected to Router A need to be transparently transmitted through the OSPF AS to the IS-IS area connected to Router D. Routes from the RIP-2 area connected to Router B need to be transparently transmitted through the OSPF AS to the RIP-2 area connected to Router C.

 Copyright © Huawei Technologies Co., Ltd.

Figure 11-176 Networking diagram for transparently transmitting routes of other protocols through an OSPF AS



To meet the preceding requirements, configure a Route-Policy for Router A to set a tag for the imported IS-IS routes. Router D identifies the IS-IS routes from OSPF routes based on the tag.

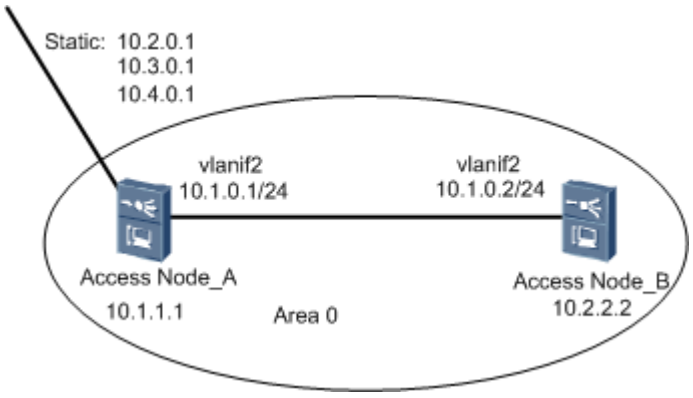
11.12.9.4 Configuration Example of the Routing Policy

This topic provides an example for configuring a routing policy for imported routes.

Service Requirements

- Consider two MA5800s with routing function enabled, namely MA5800_A and MA5800_B. Both of them are running the OSPF routing protocol, and within area 0.
- MA5800_A imports static routes, and MA5800_B is configured with the routing filtering policy.

Figure 11-177 Example network for configuring the routing policy



Procedure

Step 1 Configuring MA5800_A.

1. Configure the IP address of the Layer 3 interface on MA5800_A.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#vlan 2 smart
huawei(config)#port vlan 2 0/9 0
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.1.0.1 24
huawei(config-if-vlanif2)#quit
```

2. Enable OSPF on MA5800_A and specify the area ID to which the interface belongs.

```
huawei(config)#ospf
huawei(config-ospf-1)#area 0
huawei(config-ospf-1-area-0.0.0.0)#network 10.1.0.0 0.0.0.255
huawei(config-ospf-1-area-0.0.0.0)#quit
huawei(config-ospf-1)#quit
```

3. Configure the OSPF router ID of MA5800_A.

```
huawei(config)#router id 10.1.1.1
```

4. Configure three static routes.

```
huawei(config)#ip route-static 10.2.0.1 32 NULL 0
huawei(config)#ip route-static 10.3.0.1 32 NULL 0
huawei(config)#ip route-static 10.4.0.1 32 NULL 0
```

5. Import static routes into the OSPF routing table to improve its capability of obtaining routes.

```
huawei(config)#ospf
huawei(config-ospf-1)#import-route static
huawei(config-ospf-1)#quit
```

6. Save the data.

```
huawei(config)#save
```

Step 2 Configuring MA5800_B.

1. Configure the IP address of the Layer 3 interface on MA5800_B.

```
huawei(config)#vlan 2 smart
huawei(config)#port vlan 2 0/9 0
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.1.0.2 24
huawei(config-if-vlanif2)#quit
```

2. Configure the ACL.

```
huawei(config)#acl 2000
huawei(config-acl-basic-2000)#rule deny source 10.3.0.0 255.255.255.0
huawei(config-acl-basic-2000)#rule permit source any
huawei(config-acl-basic-2000)#quit
```

3. Enable OSPF on MA5800_B and specify the area ID to which the interface belongs.

```
huawei(config)#ospf
huawei(config-ospf-1)#area 0
huawei(config-ospf-1-area-0.0.0.0)#network 10.1.0.0 0.0.0.255
huawei(config-ospf-1-area-0.0.0.0)#quit
huawei(config-ospf-1)#quit
```

4. Configure the OSPF router ID of MA5800_B.

```
huawei(config)#router id 10.2.2.2
```

5. Configure OSPF to filter the received external routes.

```
huawei(config)#ospf
uawei(config-ospf-1)#filter-policy 2000 import
huawei(config-ospf-1)#quit
```

6. Save the data.

```
huawei(config)#save
```

----End

 Copyright © Huawei Technologies Co., Ltd.

Result

1. MA5800_A and MA5800_B run OSPF successfully, and they can communicate well with each other.
2. After a filter is configured on MA5800_B, parts of the three imported static routes are available while part of them is screened on MA5800_B. That is, routes from segments 10.2.0.0 and 10.4.0.0 are available, while the route from segment 10.3.0.0 is screened.

Configuration File

Configuration on MA5800_A.

```
vlan 2 smart
port vlan 2 0/9 0
interface vlanif 2
ip address 10.1.0.1 24
quit
ospf
area 0
network 10.1.0.0 0.0.0.255
quit
quit
router id 10.1.1.1
ip route-static 10.2.0.1 32 NULL 0
ip route-static 10.3.0.1 32 NULL 0
ip route-static 10.4.0.1 32 NULL 0
ospf
import-route static
quit
save
```

Configuration on MA5800_B.

```
vlan 2 smart
port vlan 2 0/9 0
interface vlanif 2
ip address 10.1.0.2 24
acl 2000
rule deny source 10.3.0.0 255.255.255.0
rule permit source any
quit
ospf
area 0
network 10.1.0.0 0.0.0.255
quit
quit
router id 10.2.2.2
ospf
filter-policy 2000 import
quit
save
```

11.12.9.5 Routing Policy Reference Standards and Protocols

None

 Copyright © Huawei Technologies Co., Ltd.

12 Internet Access

12.1 DHCP Emulation
In a DHCP Emulation test, an access node simulates the DHCP client to implement remote acceptance for services that obtain IP addresses in DHCP mode (such as IPTV, IPoE, and VoIP services) and to locate faults.

12.2 PPPoE Dialup Emulation

12.3 Traffic Burst Detection
Traffic burst detection is a feature of identifying traffic burst points by deploying detection points on the network and detecting the traffic at the detection points to obtain traffic information.

12.1 DHCP Emulation

In a DHCP Emulation test, an access node simulates the DHCP client to implement remote acceptance for services that obtain IP addresses in DHCP mode (such as IPTV, IPoE, and VoIP services) and to locate faults.

12.1.1 DHCP Emulation Overview

In a DHCP emulation test, an access node emulates a DHCP client to apply for an IP address from a DHCP server. By doing so, the access node can test the connectivity between the access node and DHCP server and therefore the validity of DHCP configurations on the DHCP relay agent, DHCP agent, and DHCP server. In addition, the access node can ping a remote server to test the connectivity between them.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

- During DHCP simulation, certain DHCP simulation packets can be sent in a specified mode. Improper use of some simulation fields (such as source MAC address) may cause service abnormality. Therefore, use this feature only when you understand the impact of the simulation packets.
- The access node can be an OLT, MDU, or ONT.
- The remote server is an upstream network device, for example, an IPTV server or softswitch.
- The difference between a DHCP emulation test and the actual DHCP process is the initiator. In a DHCP emulation test, the access node functions as the initiator; in the actual DHCP process, the user terminal (for example, a PC, STB, or VoIP service terminal) functions as the initiator.

Mainly used for deployment acceptance and fault locating, a DHCP emulation test resolves the known issues in the traditional method. For details, see **Table 12-1**.

Table 12-1 Comparison between the traditional method and DHCP emulation test

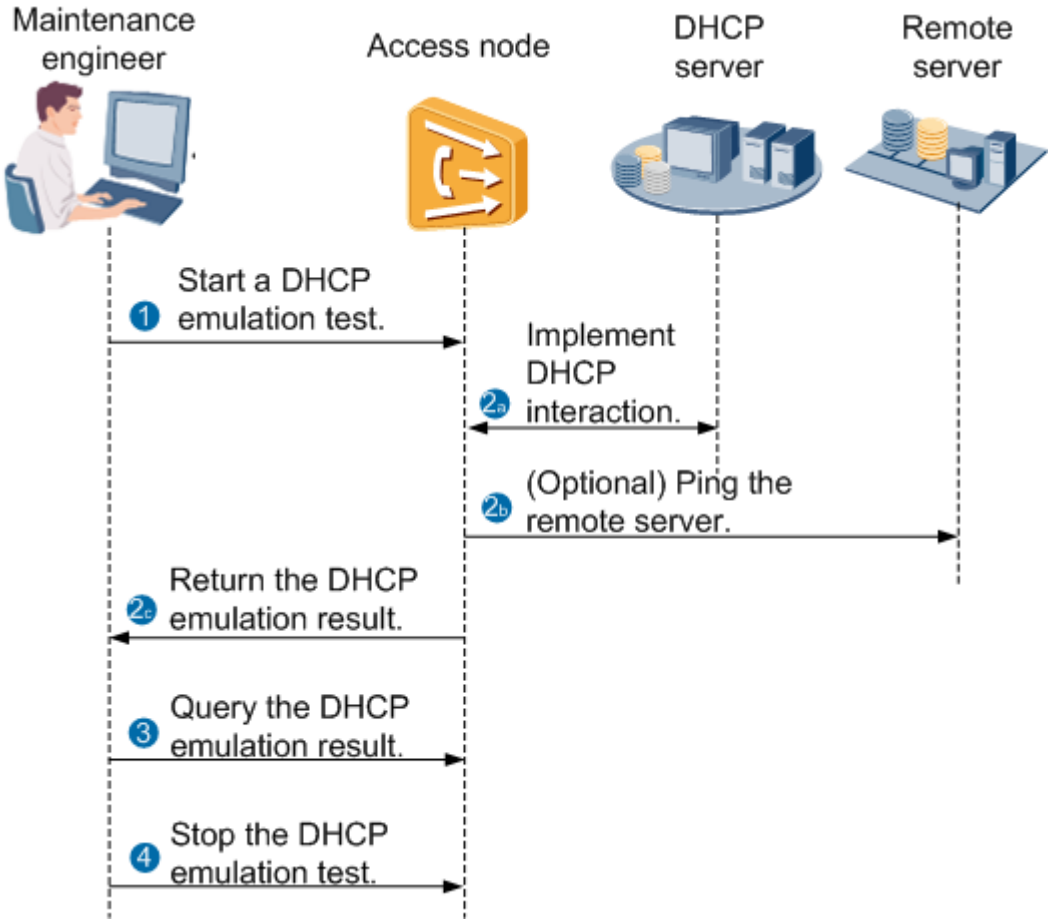
Application Scenario	Traditional Method	DHCP Emulation Test
Deployment acceptance	At the device installation site, a test instrument is connected to the device and the process of users going online in batches is emulated. Alternatively, a laptop is connected to the device to test services on each port. The O&M costs are high.	A remote node implements an acceptance test for services, which does not require site visits and reduces the O&M costs.
Fault locating	When a user reports a fault, the maintenance engineer cannot remotely locate the specific network segment in a timely manner. The maintenance engineer needs to ask for the user's cooperation or go to the user's house to locate the fault, bringing about inconvenience to the user and incurring high O&M costs.	Faults can be remotely located and the specific network segment can be quickly located. Then a work order is assigned to the corresponding maintenance engineer. In this way, the efficiency is improved and the O&M costs are reduced.

12.1.2 DHCP Emulation Principles

The principles of DHCP emulation.

 Copyright © Huawei Technologies Co., Ltd.

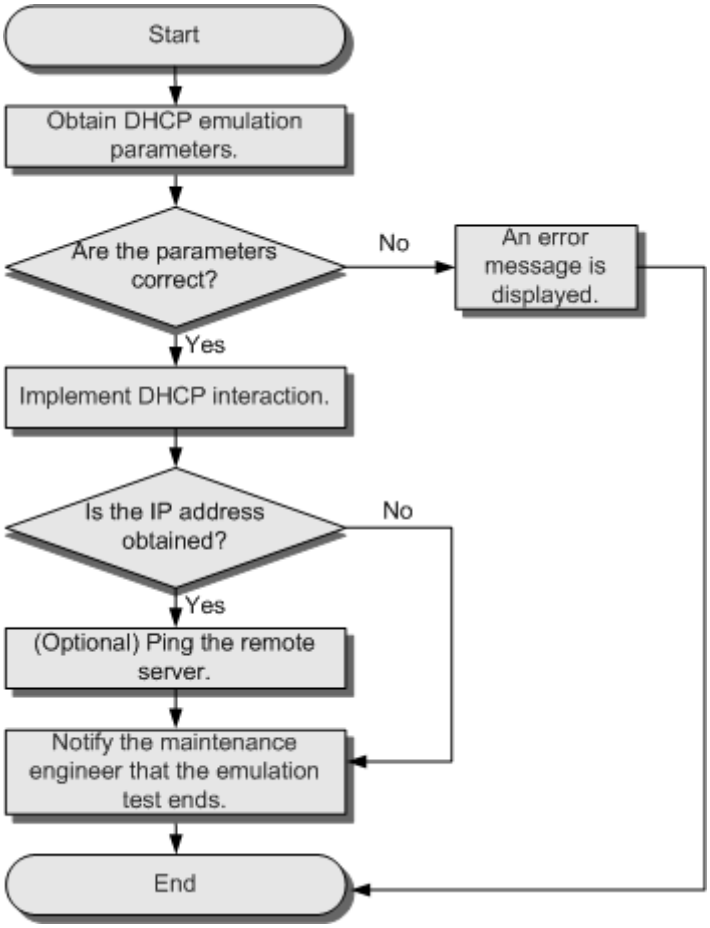
Figure 12-1 Principles of DHCP emulation



1. The maintenance engineer remotely logs in to the access node and starts a DHCP emulation test.
2. The DHCP emulation process starts. **Figure 12-2** shows the DHCP emulation process.
   a. DHCP interaction is implemented between the access node and DHCP server, which tests the link connectivity between them.
   b. After obtaining an IP address, the access node pings the remote server if **hostip** (a DHCP emulation parameter indicating the IP address of the remote server) is set. This tests the link connectivity between the access node and remote server.
   c. The DHCP emulation is completed and the DHCP emulation result is returned.

 Copyright © Huawei Technologies Co., Ltd.

Figure 12-2 DHCP emulation flow



3. The maintenance engineer queries the DHCP emulation result.
4. The maintenance engineer manually stops the DHCP emulation test to release system resources in a timely manner.

NOTE

A DHCP emulation test automatically stops when the default timeout interval (210s) elapses and the emulation result will be deleted.

12.1.3 DHCP Emulation Usage Scenario

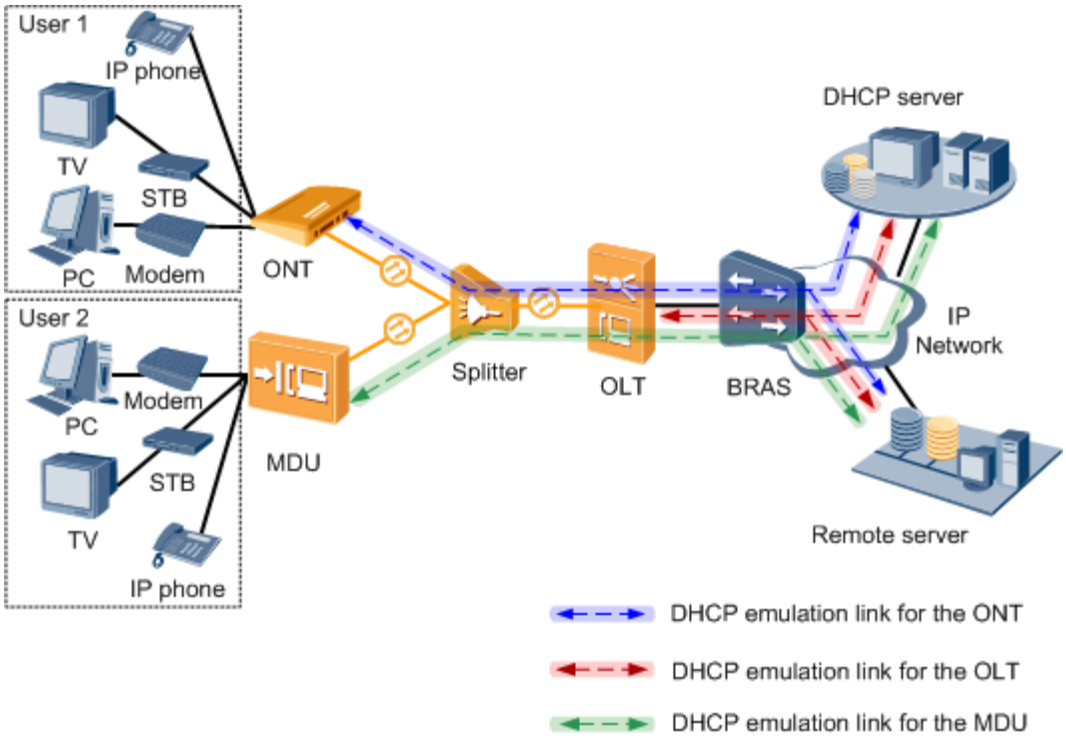
On an FTTx network, access nodes are close to user terminals and are widely deployed. In a DHCP emulation test, an access node emulates the DHCP client to implement remote acceptance for services that obtain IP addresses in DHCP mode (such as IPTV, IPoE, and VoIP services) and locate faults, which reduces the O&M costs.

FTTx Network Application Scenarios

Figure 12-3 shows typical application scenarios of DHCP emulation on an FTTx network.

 Copyright © Huawei Technologies Co., Ltd.

Figure 12-3 Typical application scenarios of DHCP emulation on an FTTx network



NOTE

In **Figure 12-3**:

- Users 1 and 2 receive IP services (such as IPTV, IPoE, and VoIP services) and are authenticated through the BRAS.
- User 1, an FTTH user, receives IP services through the ONT; user 2, an FTTB/FTTC user, receives IP services through the MDU.

When a user receives IPTV, IPoE, or VoIP services and implements deployment acceptance or fault locating through a DHCP emulation test, only the user terminals and remoter servers differ and the operation methods are the same. The following uses receiving IPTV services as an example.

When IPTV services are undergoing deployment acceptance or faulty:

- If the STB cannot obtain an IP address, network connectivity between the STB and DHCP server is abnormal. Perform a DHCP emulation test on the access node device to test the connectivity of each network segment.
- If the STB obtains an IP address successfully, network connectivity between the STB and DHCP server is normal. If users 1 and 2 still cannot watch IPTV programs, ping the remote server (for example, an IPTV server) to test the connectivity between the STB and remote server.

Fault Locating

Locate the fault based on the obtained DHCP emulation result, as shown in **Table 12-2**.

 Copyright © Huawei Technologies Co., Ltd.

Table 12-2 Mapping between the emulation result and fault range

emulation Result	Description	Fault Scope
send packet fail	The access node failed to send packets to the DHCP server.	The access node is faulty.
DHCP server not found	After sending DHCP Discover packets, the access node did not receive response packets from the DHCP server.	The network segment between the access node and DHCP server is faulty.
DHCP server refuse request	The DHCP server rejected the DHCP request from the access node.	The DHCP server is faulty.
get IP overtime	After sending a DHCP request, the access node did not receive confirmation packets from the DHCP server.	The network segment between the access node and DHCP server is faulty.
get IP successful	The access node obtained an IP address from the DHCP server, indicating that the DHCP interaction was successful.	The possible fault scopes are as follows: • The network segment between the user terminal and access node is faulty. • The network segment between the BRAS and remote server is faulty.
all ping succeed	The access node succeeded in pinging all the specified remote servers.	The network segment between the user terminal and access node is faulty.
partial ping succeed	The access node succeeded in pinging some of the specified remote servers.	The network segment between the BRAS and remote server that failed to be pinged is faulty.
all ping fail	The access node failed to ping all the specified remote servers.	Network segments between the BRAS and all remote servers are faulty.
system resource not enough	On the access node, resources for DHCP emulation are insufficient.	On the access node, resources for DHCP emulation are insufficient.

 Copyright © Huawei Technologies Co., Ltd.

emulation Result	Description	Fault Scope
unknown	-	-

12.1.4 Configuring DHCP Emulation

Context

Maintenance engineers can log in to the access node and perform DHCP emulation configurations through the CLI or NMS. The following uses operations through the CLI as an example.

- For FTTB or FTTC scenario, a DHCP emulation test can be directly performed on the OLT or MDU.
- For FTTH scenario, a DHCP emulation test can be started on the ONT through the OLT.

Procedure

Step 1 Run the **simulate dhcp start** command to start a DHCP emulation test.

Step 2 Run the **display simulation dhcp** command to query the specified DHCP emulation result.

NOTE

When the default timeout interval (210s) for the DHCP emulation test elapses or the DHCP emulation test is manually stopped, the DHCP emulation result will be deleted. Therefore, after the DHCP emulation test is completed, query the DHCP emulation result in a timely manner.

Locate the fault based on the obtained DHCP emulation result, as shown in **Table 12-2**.

Step 3 Run the **simulate dhcp stop** command to stop the DHCP emulation test.

----End

Example

In the FTTB or FTTC scenario: Log in to the OLT or MDU and start a DHCP emulation test on it.

- ID of the service port: 0
- MAC address of the simulated server: 00e0-fc00-0001
- Option 60: HW
- Option 61 type corresponding to the option 61 field: 1; client ID: 00e0-fc00-0002
- IP address of the to-be-pinged remote server: 192.168.3.99

1. Start a DHCP emulation test.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#simulate dhcp start service-port
{ service-portid<U><0,131071> }:0
{ mac-address<P><H-H-H>  }:00e0-fc00-0001
{ <cr>|host-ip1<K>|host-ip2<K>|host-ip3<K>|host-ip4<K>|host-ip5<K>|option60<K>|o
ption61<K>|user-vlan<K> }:option60
{ option60<S><Length 1-64> }:HW
{ <cr>|host-ip1<K>|host-ip2<K>|host-ip3<K>|host-ip4<K>|host-ip5<K>|option61<K> }
:option61
{ mac-addr<K>|string<K> }:mac-addr
{ mac_clientid<P><XXXX-XXXX-XXXX> }:00e0-fc00-0002
{ <cr>|host-ip1<K>|host-ip2<K>|host-ip3<K>|host-ip4<K>|host-ip5<K> }:host-ip1
{ ipaddr1<I><X.X.X.X> }:192.168.3.99
{ <cr>|host-ip2<K>|host-ip3<K>|host-ip4<K>|host-ip5<K>|overtime<K> }:

 Command:
        simulate dhcp start service-port 0 00e0-fc00-0001 option60 HW
option61 mac-addr 00e0-fc00-0002 host-ip1 192.168.3.99

 Please check whether the input flow ID and MAC address are correct. Incorrect
parameters may affect the services of normal users.
 You are suggested to use the MAC address of the actual user.
 Are you sure to continue?(y/n)[n]:y
```

2. Query the DHCP emulation result.

```
huawei(config)#display simulation dhcp
{ <cr>|service-port<K>||<K>|><K>|>><K> }:service-port
{ service-portid<U><0,131071> }:0
{ mac-address<P><XXXX-XXXX-XXXX> }:00e0-fc00-0001

 Command:
        display simulation dhcp service-port 0 00e0-fc00-0001
 Total : 1
 ------------------------------------------------------------------------------
 --------------------------------- common info ---------------------------------
 FlowID                 : 0
 User MAC               : 00e0-fc00-0001
 User VLAN              : 10
 Optin60                : HW
 Option61 type          : 1
 Option61 client        : 00e0-fc00-0002
 Simulate status        : finished
 Simulate result        : all ping succeed
 Simulate start time    : 2019-01-17 16:51:59+08:00
 User IP                : 192.168.3.100/24
 DHCP server IP         : 192.168.3.8
 Primary DNS server IP  : -
 Secondary DNS server IP : -
 Relay gateway          : 0.0.0.0
 Default gateway        : 192.168.3.201
 --------------------------------- route info ---------------------------------
 Route:1  Destination IP : 192.168.5.0/24     Next hop IP : 192.168.3.66
        Next hop MAC   : -                Local interface : -
 Route:2  Destination IP : 192.168.4.0/24     Next hop IP : 192.168.3.88
        Next hop MAC   : 00e0-fc00-dbba     Local interface : Y
 Route:3  Dstination IP  : -                Next hop IP : -
        Next hop MAC   : -                Local interface : -
 ---------------------------------- ping info ----------------------------------
 Host:1   IP : 192.168.3.99       Total count : 3     Succeed count : 3
 Host:2   IP : -                  Total count : 0     Succeed count : 0
 Host:3   IP : -                  Total count : 0     Succeed count : 0
 Host:4   IP : -                  Total count : 0     Succeed count : 0
 Host:5   IP : -                  Total count : 0     Succeed count : 0
 ------------------------------------------------------------------------------
```

3. Stop the DHCP emulation test.

```
huawei(config)#simulate dhcp stop service-port 0 00e0-fc00-0001
```

In the FTTH scenario, log in to the OLT and start a DHCP emulation test on the ONT.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

During the DHCP emulation, the IP address obtained can be used for multicast emulation.

- ID of the subrack/slot/port of the OLT connected to the ONT: 0/2/0
- ID of the ONT enabled with DHCP emulation: 1
- VLAN ID in the VLAN tag in packets: 10
- Option 61 type corresponding to the option 61 field: 0; client ID: user
- IP address of the to-be-pinged remote server: 192.168.3.99
- IP address of the multicast program: 224.1.1.1

1. Issue commands on the OLT to start a DHCP emulation test on the ONT.

```
huawei(config)#simulate dhcp start ont
{ frameid/slotid/portid<S><Length 5-18> }:0/2/0
{ ontid<U><0,255> }:1
{ eth<K>|outer-vlan<K>|untag<K>|vlan<K> }:vlan
{ vlanid<U><1,4095> }:10
{ <cr>|host-ip1<K>|host-ip2<K>|host-ip3<K>|host-ip4<K>|host-ip5<K>|option60<K>|o
ption61<K>|priority1<U><0,7>|usermac<K> }:option61
{ mac-addr<K>|string<K> }:string
{ string_clientid<S><Length 1-64> }:user
{ <cr>|host-ip1<K>|host-ip2<K>|host-ip3<K>|host-ip4<K>|host-ip5<K> }:host-ip1
{ ipaddr1<I><X.X.X.X> }:192.168.3.99
{ <cr>|host-ip2<K>|host-ip3<K>|host-ip4<K>|host-ip5<K>|groupid<K> }:groupid
{ip-addr<I><X.X.X.X>|<cr> }:224.1.1.1
{igmp-version<K> }:igmp-version
{v2<K>|v3<K>}:v2

 Command:
        simulate dhcp start ont 0/2/0 1 vlan 10 option61 string user host-ip1
 192.168.3.99 groupid 224.1.1.1 igmp-version v2
```

2. Query the ONT DHCP emulation result.

```
huawei#display simulation dhcp
{ <cr>|ont<K>|service-port<K>||<K>|><K>|>><K> }:ont
{ <cr>|frameid/slotid/portid<S><Length 5-18> }:0/2/0
{ ontid<U><0,255> } 1

 Command:
        display simulation dhcp ont 0/2/0 1
------------------------------------------------------------------------------
-------------------------------- common info --------------------------------
simulation instance     : F/S/P: 0/2/0  ONTId:1  ONTPort WAN
simulation VLAN tag num : 1
simulation outer VLANID : 10
simulation inner VLANID : -
Option60                : -
Option61 type           : 0
Option61 client         : user
Simulate status         : finished
Simulate result         : succeed
Simulate start time     : 2013-01-17 15:49:20+09:00
User IP                 : 192.168.3.100/24
DHCP server IP          : 192.168.3.99
Primary DNS server IP   : 10.72.55.81
Secondary DNS server IP : 10.72.255.100
Relay gateway           : 192.168.3.99
Default gateway         : 192.168.3.99
-------------------------------- route info --------------------------------
Route:1  Destination IP : 192.168.5.0/24     Next hop IP : 192.168.3.66
        Next hop MAC    : -                  Local interface : -
Route:2  Destination IP : 192.168.4.0/24     Next hop IP : 192.168.3.88
        Next hop MAC    : 00e0-fc00-dbba     Local interface : Y
Route:3  Dstination IP  : -                  Next hop IP : -
        Next hop MAC    : -                  Local interface : -
-------------------------------- ping info --------------------------------
```

 Copyright © Huawei Technologies Co., Ltd.

```
Host:1    IP : 192.168.3.99       Total count : 3     Succeed count : 3
Host:2    IP : -                  Total count : 0     Succeed count : 0
Host:3    IP : -                  Total count : 0     Succeed count : 0
Host:4    IP : -                  Total count : 0     Succeed count : 0
Host:5    IP : -                  Total count : 0     Succeed count : 0
--------------------------------------------------------------------------------
--------------------------------------------------------------------------------
 Multicast group IP      : 224.1.1.1
 Multicast source IP     : -
 Multicast version       : IGMPv2
 Data rate(kbps)         : 1500
--------------------------------------------------------------------------------
```

3. Stop the ONT DHCP emulation test.

```
huawei(config)#simulate dhcp stop ont 0/2/0 1
```

12.1.5 DHCP Emulation Reference Standards and Protocols

- IETF RFC 2131: Dynamic Host Configuration Protocol
- IETF RFC 1533: DHCP Options and BOOTP Vendor Extensions

12.2 PPPoE Dialup Emulation

12.2.1 PPPoE Dialup Emulation Introduction

In PPPoE dialup emulation, an access device emulates an end user and initiates PPPoE dialup. According to the PPPoE dialup emulation result (successful or failed), users can determine the network connectivity between the access device and the broadband remote access server (BRAS). If a fault has occurred in the network, PPPoE dialup emulation helps users quickly identify whether the fault is on the network side or user side.

NOTE

During PPPoE simulation, certain PPPoE simulation packets can be sent in a specified mode. Improper use of some simulation fields (such as source MAC address) may cause service abnormality. Therefore, use this feature only when you understand the impact of the simulation packets.

PPPoE dialup emulation is used for remote fault location and acceptance tests. **Table 12-3** lists the comparison between the method of performing a PPPoE dialup emulation test and the method of performing a traditional PPPoE dialup test.

 Copyright © Huawei Technologies Co., Ltd.

Table 12-3 Comparison between the method of performing a PPPoE dialup emulation test and the method of performing a traditional PPPoE dialup test

Scenario	Task	Method of Performing a Traditional PPPoE Dialup Test	Method of Performing a PPPoE Dialup Emulation Test
Acceptance test	After an access device is installed, a test engineer needs to check whether the PPPoE dialup service on the access device has been provisioned properly.	The test engineer visits the site where the access device is installed and uses an external tester or a portable computer to perform a PPPoE dialup test for each port.	The test engineer remotely logs in to the access device to perform a PPPoE dialup emulation test and determines the service status based on the test result. **NOTE** A PPPoE dialup emulation test cannot check the status of the line between an end user and an access device.
Fault location	A maintenance engineer needs to locate a fault on an access network that covers a large area and contains geographically dispersed network devices.	The maintenance engineer visits all sites where the access devices are installed and performs a PPPoE dialup test.	The maintenance engineer preliminarily determines the network segment where the fault occurs and remotely logs in to the access devices to perform a PPPoE dialup emulation test. Based on the test result, the maintenance engineer determines the fault cause and rectifies the fault.

12.2.2 PPPoE Dialup Emulation Principles

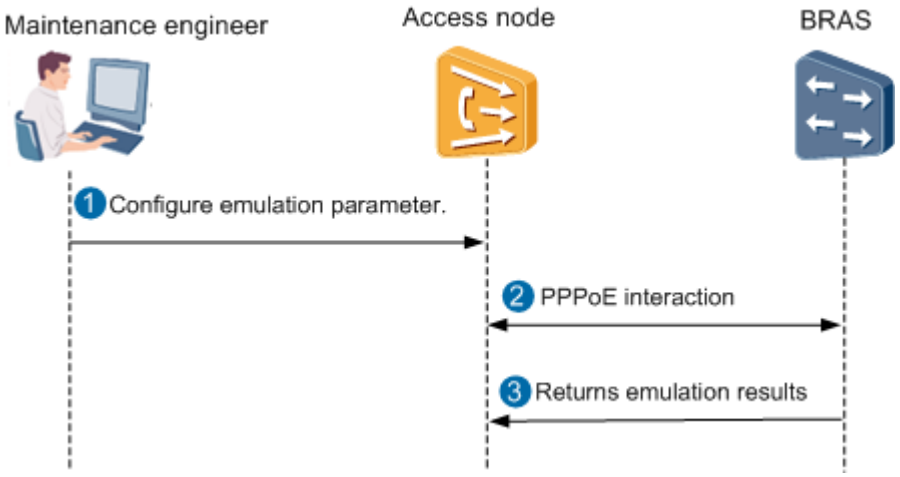
PPPoE dialup emulation differs from PPPoE dialup in the following aspects:

- The PPPoE dialup emulation process is initiated by an access device, while the PPPoE dialup is initiated by a PC, modem, or home gateway.
- The MAC address carried in the packet in PPPoE dialup emulation is the bridge MAC of an access device, while the MAC address carried in the packet in the PPPoE dialup is that of a PC, modem, or home gateway.

The principles of PPPoE dialup emulation.

 Copyright © Huawei Technologies Co., Ltd.

Figure 12-4 Principles of PPPoE dialup emulation



1. A maintenance engineer can start a PPPoE dialup emulation on an access node, the access node emulates an end user and initiates PPPoE dialup.

   Enter parameters required for the PPPoE dialup emulation, including service flow ID (identifying the user), MAC address, user name, password, authentication mode, and timeout time.

2. Access node interacts with the BRAS.

   The emulated PPPoE dialup includes two phases:

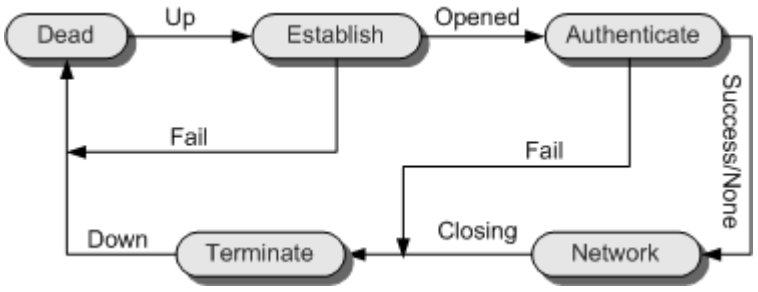
   a. PPPoE discovery phase.

      The source host discovers the MAC address of the destination host (in this case, the BRAS), and a point-to-point PPP session is set up.

   b. PPPoE session phase.

      A point-to-point connection has been set up and carries PPP packets for link negotiation, as shown in **Figure 12-5**.

      Figure 12-5 PPPoE session flow

3. The maintenance engineer can stop a PPPoE dialup emulation on an access node.

 Copyright © Huawei Technologies Co., Ltd.

12.2.3 PPPoE Dialup Emulation Usage Scenario

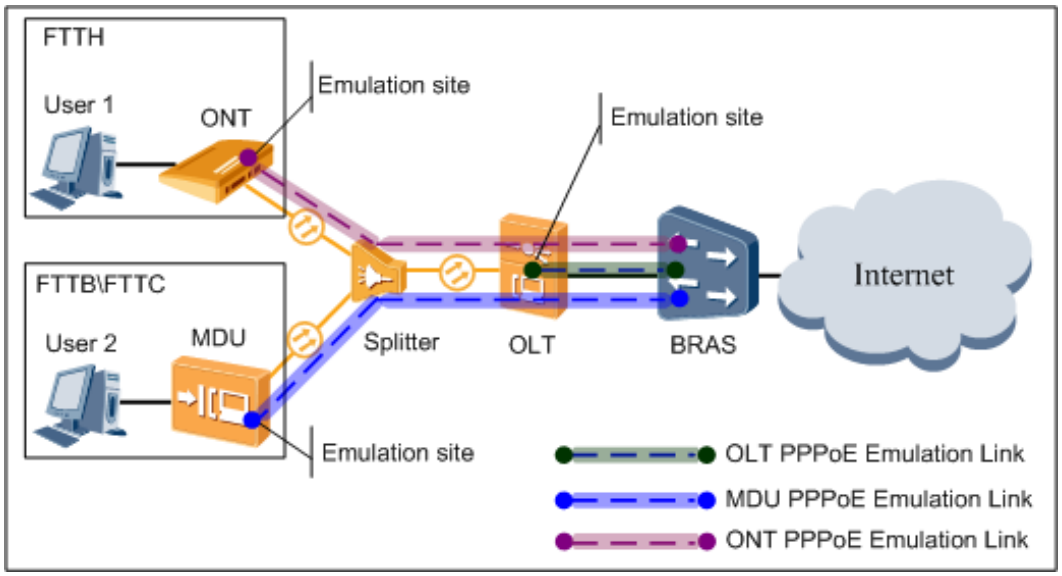
Context

On networks, access devices are closer to terminals and widely distributed. If a PPPoE dialup fault occurs, PPPoE dialup tests performed onsite increase fault location costs. To resolve this issue, PPPoE dialup emulation tests can be performed remotely.

Scenario

Figure 12-6 shows the typical networking.

Figure 12-6 Typical networking



In the preceding figure, users use the PPPoE dialup service and they are authenticated in PPPoE mode through the broadband remote access server (BRAS).

- For ONT user, perform the PPPoE dialup emulation on the ONT.
- For MDU user, perform the PPPoE dialup emulation on the MDU.
- For OLT user, perform the PPPoE dialup emulation on the OLT.

Fault Location

The troubleshooting roadmap based on the test result is as follows:

 Copyright © Huawei Technologies Co., Ltd.

Table 12-4 PPPoE dialup emulation results

Emulation Result	Corresponding Windows Error Code	Description	Handling Guide
Success	-	• The link between the access node and the BRAS is functional. • The service configuration of the PPPoE user is correct. • Parameters such as the user name, password, and authentication mode of the PPPoE user are correct.	-
Parameter negotiation fail	732 Your computer and the remote computer could not agree on PPP control protocols.	PPPoE dialup emulation has entered the link setup phase of the PPP session, but parameter negotiation fails, causing the link failure.	Check whether the Policy Information Transfer Protocol (PITP) configuration is correct.
Link unnormal	638 The remote server is not responding in a timely fashion.	PPPoE dialup emulation has entered the link setup phase of the PPP session, but the creation of the LCP link is abnormal.	Check whether an access control list (ACL) that does not allow transmission of PPPoE packets is configured on the access node.

 Copyright © Huawei Technologies Co., Ltd.

Emulation Result	Corresponding Windows Error Code	Description	Handling Guide
Authentication fail	691 The connection was denied because the username or password you specified is not valid or because the selected authentication protocol is not permitted on the remote server.	PPPoE dialup emulation has entered the user authentication phase of the PPP session, but user name or password mismatch occurs between the client and server.	• The user's account is configured on the access node is not correct. • The user's account is configured on the BRAS is not correct.
Time out	721 The remote computer is not responding.	PPPoE dialup emulation exceeds the preset emulation timeout time.	• The present emulation time out time is too short. • The link between access node and BRAS is abnormal, transmission delay is too large.
Peer down request	668 The connection was terminated.	The server forcibly terminates the PPPoE dailup emulation.	The user's account is restricted on the BRAS.
Other error	635 There was an unknown error.	Unknown errors.	-

12.2.4 Configuring PPPoE Dialup Emulation

Context

If a PPPoE dialup fault occurs, a PPPoE dialup emulation test can be performed on the ONT or MDU to check the connection between the ONT or MDU and the BRAS and locate the fault according to the test result.

Procedure

Step 1 Run the **simulate pppoe start** command to perform a PPPoE dialup emulation.

Step 2 Run the **display simulatation pppoe** command to query the status of the PPPoE dialup emulation.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

Query the status of the PPPoE dialup emulation only after the test is started. The status cannot be queried if the emulation is stopped.

Step 3 Run the **simulate pppoe stop** command to stop PPPoE dialup emulation.

NOTE

Another PPPoE dialup emulation can be performed only after the current PPPoE dialup emulation is stopped.

----End

Example

In FTTB or FTTC scenario, remotely log in to the access node device and perform PPPoE dialup emulation on it.

- Service port ID: 0
- MAC address of the emulation user: 00e0-fc00-1111

1. Start PPPoE dialup emulation.
```
huawei(config)#simulate pppoe start
 Service-port(index<0-1999>):0
 Mac-address<P><XXXX-XXXX-XXXX>[default 5623-5987-dead]:00e0-fc00-1111
 User Name(length<1,65>):user-0
 User Password(length<0,16>):*******
 Authentication Mode:
 1. Chap  2. Pap [default 1]:1
 Overtime Time(5-60s)[default 5]:10
```
2. Query the status of PPPoE dialup emulation.
```
huawei(config)#display simulation pppoe
  PPPoE simulate information is:
  ------------------------------------------------------------------
  Service-port:  0
  Mac-address:   00e0-fc00-1111
  User name:     user-0
  Current phase: -
  Result:        Success
  Start time:    2019-03-06 08:20:23+08:00
  End time:      2019-03-06 08:23:36+08:00
  Session ID:    591
  User IP:       192.168.1.172
  Gateway IP:    192.168.1.1
  ------------------------------------------------------------------
```
3. Stop PPPoE dialup emulation.
```
huawei(config)#simulate pppoe stop
```

In FTTH scenario, remotely log in to the OLT and perform PPPoE dialup emulation on the ONT.

NOTE

During the PPPoE dialup emulation, the IP address obtained can be used for multicast emulation.

- Port to which the ONT connects: 0/2/0
- ONT ID: 0
- Ethernet port: 1
- VLAN ID: 100

 Copyright © Huawei Technologies Co., Ltd.

- IP address of the multicast program: 225.0.0.1
- IP address of the multicast source: 10.2.3.4

1. Start PPPoE dialup emulation.

```
huawei(config-if-gpon-0/2)#pppoe simulate start
{ portid<U><0,3>}:0
{ ontid<U><0,63>}:0
{ eth<K>|untagged<K>|vdsl<K>|vlanid<U><0,4095> }:eth
{ ont-portid<U><1,24>}:1
{ untagged<K>|vlanid<K> }:100
{ priority<U><0,7>|user-name<K> }:2
{user-name<K> }:user-name
{user-name<S><Length 1-64> }:pppoe
{ user-password<K> }:user-password
{ password<S><Length 1-64> }:password
{ authentication-mode<K> }:authentication-mode
{ protocol<U><chap,pap> }:chap
{ <cr>|group-ip-address<K> }:group-ip-address
{ group-ip-address<I><X.X..X> }:225.0.0.1
{ <cr>|igmp-version<K> }:igmp-version
{v2<K>|v3<K>}:v3
{ <cr>|source-ip-address<K> }:source-ip-address
{ source-ip-address<I><X.X..X> }:10.2.3.4

Command:
       pppoe simulate start 0 0 eth 1 100 2 user-name pppoe user-password password authentication-mode chap groupip 225.0.0.1 igmp-version v2 source-ip-address 10.2.3.4
  ---------------------------------------------------------------------------
  ONT PPPoE Test Result
  ---------------------------------------------------------------------------
  F/S/P                  : 0/2
  ONT-ID                 : 0
  ONT ETH Port ID        : 1
  ONT Vlan ID            : 100
  Vlan Priority          : 2
  Multicast group IP     : 225.0.0.1
  Multicast source IP    : 10.2.3.4
  Multicast version      : v2
  Multicast traffic(kbps) : 1500
  Emluator result        : Multicast test fail
  Session ID             : 0
  User IP                : 192.168.1.1
  Gateway IP             : 255.255.255.0
---------------------------------------------------------------------------
```

2. Query ONT PPPoE dialup emulation results of port 0/2.

```
huawei(config-if-gpon-0/2)#display pppoe simulate 0
{<cr>|ontid<U><0,127>}

 Command:
        display pppoe simulate 0
---------------------------------------------------------------------------
  F/S/P                  : 0/2
  ONT-ID                 : 0
  ONT ETH Port ID        : 1
  ONT Vlan ID            : 100
  Vlan Priority          : 2
  Multicast group IP     : 225.0.0.1
  Multicast source IP    : 10.2.3.4
  Multicast version      : v2
  Multicast Data rate(kbps): 1500
  Emluator result        : Multicast test fail
  Session ID             : 0
  User IP                : 192.168.1.1
  Gateway IP             : 255.255.255.0
```

3. Run the **simulate pppoe stop** command to stop PPPoE dialup emulation.

```
huawei(config)#simulate pppoe stop
```

 Copyright © Huawei Technologies Co., Ltd.

12.2.5 PPPoE Dialup Emulation Reference Standards and Protocols

IETF RFC 2516: A Method for Transmitting PPP Over Ethernet (PPPoE)

IETF RFC 1661: The Point-to-Point Protocol (PPP)

12.3 Traffic Burst Detection

Traffic burst detection is a feature of identifying traffic burst points by deploying detection points on the network and detecting the traffic at the detection points to obtain traffic information.

12.3.1 Traffic Burst Detection Overview

In the high bandwidth era, traffic burst occurs irregularly in peak hours of network services, causing congestion or committed access rate (CAR)-triggered packet loss and affecting the performance of the end-to-end (E2E) network. Therefore, it is important to identify traffic burst points on the network.

Conventional detection methods are complex and time-consuming and do not apply to Gigabit Ethernet (GE)-level traffic burst.

Traffic burst detection enables you to easily monitor gigabit-level traffic burst on the E2E network and identify traffic burst points.

12.3.2 Traffic Burst Detection Principle

Traffic burst detection enables you to monitor traffic status in the detection period and determine whether traffic burst occurs in the detection period. Traffic burst detection can be performed based on either of the following factors:

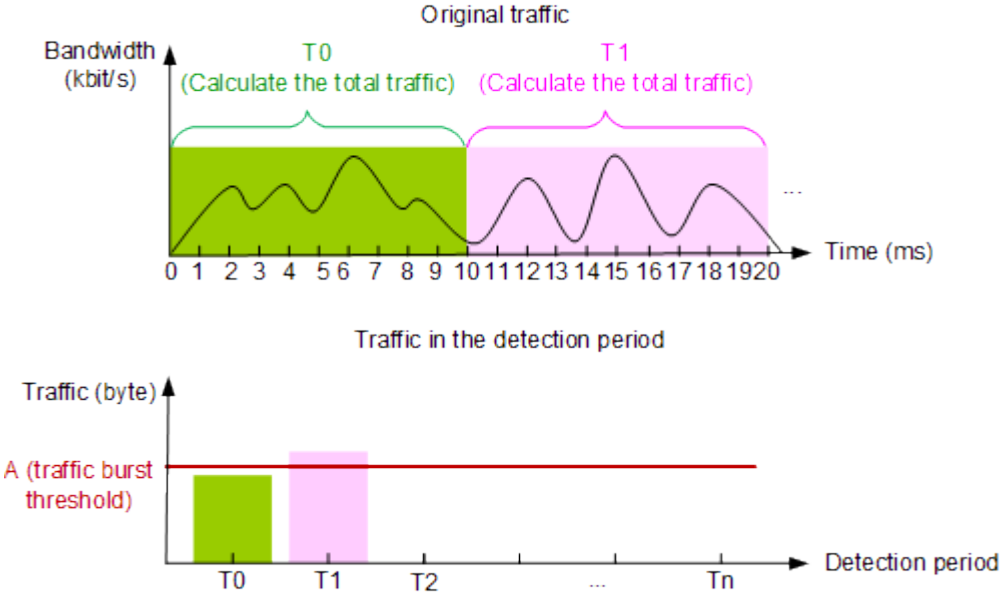
- Periodic traffic
- Rate limitation threshold

Traffic Burst Detection Based on Periodic Traffic

Figure 1 shows the operating principle of traffic burst detection based on periodic traffic.

 Copyright © Huawei Technologies Co., Ltd.

Figure 12-7 Operating principle of traffic burst detection based on periodic traffic



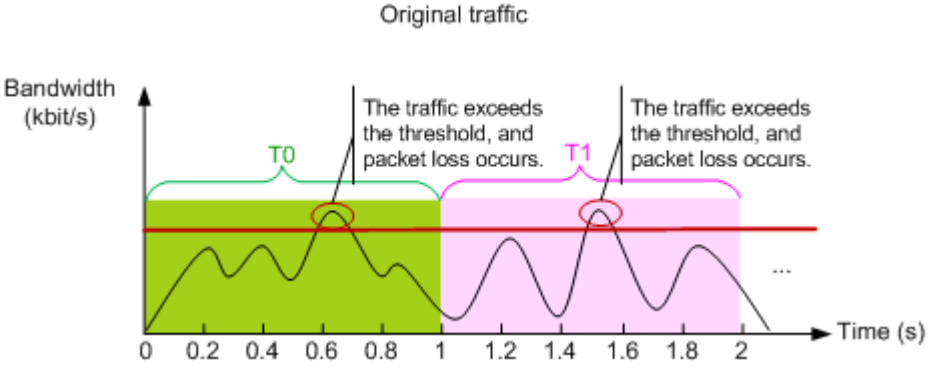
In this scenario, the traffic in each detection period (for example, T = 10 ms) is calculated, and the system determines whether the traffic in each detection period exceeds the threshold.

- If the traffic in the detection period (for example, T0) does not exceed the threshold, the system checks the traffic in the next detection period (for example, T1).
- If the traffic in the detection period (for example, T1) exceeds the threshold, the system generates a traffic burst record and checks the traffic in the next detection period (for example, T2).

Traffic Burst Detection Based on the Rate Limitation Threshold

Figure 2 shows the operating principle of traffic burst detection based on the rate limitation threshold.

Figure 12-8 Operating principle of traffic burst detection based on the rate limitation threshold



In this scenario, a virtual rate limitation threshold (not actually taking effect) is set for the to-be-detected traffic. The system determines whether traffic burst occurs

 Copyright © Huawei Technologies Co., Ltd.

in each detection period (for example, T = 1s) based on whether virtual packet loss is identified.

- If no packet loss is identified, no traffic burst occurs, and the system checks the traffic in the next detection period.
- If packet loss is identified, traffic burst occurs, and the system generates a traffic burst record and checks the traffic in the next detection period.

NOTE

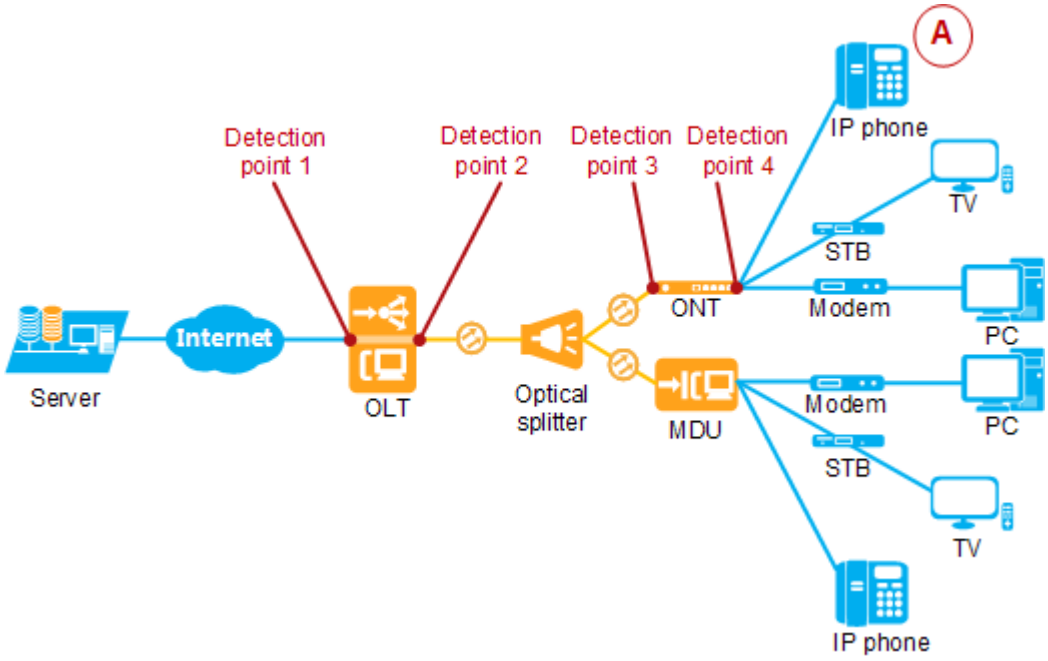
In traffic burst detection based on the rate limitation threshold, the system determines only whether traffic burst occurs, but cannot identify the specific traffic burst value.

12.3.3 Traffic Burst Detection Application Scenario

Scenario Description

Figure 1 shows the application scenario of traffic burst detection on the network.

Figure 12-9 Application scenario of traffic burst detection



Traffic burst detection is performed on different ports of the devices, to identify the traffic burst points on the network. Table 1 describes the detection points and detection directions supported on the devices.

Table 12-5 Detection points and detection directions on different devices

Device	Detection Point	Detection Direction
OLT	Upstream port(Detection point 1)	Upstream and downstream

 Copyright © Huawei Technologies Co., Ltd.

Device	Detection Point	Detection Direction
	User port(Detection point 2)	Upstream and downstream
ONT	WAN port (Detection point 3)	Upstream and downstream
	User port(Detection point 4)	Upstream and downstream

NOTE

Traffic burst detection is not limited to specific structures of the upstream network of the optical line terminal (OLT) or the user network.

Detection Point Deployment

Scenario description: In the upstream direction, traffic burst occurs on an intermediate device, causing congestion- or committed access rate (CAR)-triggered packet loss on the device at point A.

Detection point deployment: Deploy detection points at the detection point 1 and detection point 2 and at the detection point 4, to identify the device where traffic burst occurs. Functions of the detection points are as follows:

- The detection point 1 is used to identify whether traffic burst occurs on the upstream network or occurs on the OLT and its downstream network.
- The detection point 2 is used to identify whether traffic burst occurs on the OLT and its upstream network or occurs on the ONT and its downstream network.
- The detection point 4 is used to identify whether traffic burst occurs on the ONT and its upstream network or occurs on the user network.

NOTE

Traffic burst detection points in the egress direction of the ports do not cover queue scheduling. Traffic burst caused by queue scheduling in the egress direction can be identified only on the downstream device, instead of on this device.

12.3.4 Configure Traffic Burst Detection

Procedure

Step 1 Run the **traffic-burst-detect instance** command to configure traffic burst detection instances.

NOTE

You can configure instances for the optical network terminal (ONT) after logging in to the optical line terminal (OLT).

Step 2 Run the **traffic-burst-detect start** command to modify the start/end time of traffic burst detection instances.

 Copyright © Huawei Technologies Co., Ltd.

Step 3 During traffic burst detection, run the **traffic-burst-detect stop** command to stop the detection as required.

Step 4 Query traffic burst detection instances.

- Run the **display traffic-burst-detect instance all** command to query all configured traffic burst detection instances in batches.
- Run the **display traffic-burst-detect instance instance-id** command to query a single traffic burst detection instance.

Step 5 Run the **display traffic-burst-detect result** command to query traffic burst detection records.

Step 6 Run the **undo traffic-burst-detect** command to delete traffic burst detection instances.

NOTE

In a service stream, delete the original traffic burst detection instance when you attempt to configure a new instance in the same direction at a port.

----End

Example

Provided that the service stream ID is 11 and a traffic burst detection instance needs to be configured at the downstream input port 0/9/0, detailed parameter settings are as follows:

- Access control list (ACL) rule ID of the service stream: 2000
- ACL filtering rule ID of the service stream: 5
- Traffic burst detection period (ms): 10
- Traffic burst detection threshold (kbit/s): 1000

1. Configure the traffic burst detection instance.

```
huawei(config)#traffic-burst-detect instance 11
{ inbound<K>|outbound<K> }:inbound
{ ip-group<K>|ipv6<K>|link-group<K>|period<K>|user-group<K> }:ip-group
{ access-list-number2<U><2000,3999> }:2000
{ rule<K> }:rule
{ rule-id<U><0,4294967294> }:5
{ link-group<K>|period<K> }:period
{ period<U><1,100> }:1
{ threshold<K> }:threshold
{ threshold<U><64,10240000> }:1000
{ link-aggregation<K>|ont<K>|port<K>|protect-group<K> }:port
{ frameid/slotid/portid<S><Length 5-18> }:0/9/0
{ <cr>|detect-time<K>|from<K>|to<K> }:

Command:
        traffic-burst-detect  instance  11 inbound  ip-group  2000 rule  5 period 10 threshold  1000
port 0/9/0
```

2. Modify the start/end time of the traffic burst detection instance.

```
huawei(config)#traffic-burst-detect start instance 11
{<cr>|detect-time<K>|from<K>|to<K> }:from
{ time1<T><hh:mm:ss> }:10:03:10
{data1<D><yyyy-mm-dd> }: 2019-09-15
{<cr>|detect-time<K>|to<K> }:to
{ time2<T><hh:mm:ss> }:10:04:10
{data2<D><yyyy-mm-dd> }:2019-09-15
```

 Copyright © Huawei Technologies Co., Ltd.

```
Command:
        traffic-burst-detect start instance 11 from  10:03:10 2019-09-15 to 10:04:10 2019-09-15
```

3. Stop the traffic burst detection instance.

```
huawei(config)#traffic-burst-detect stop instance 11
```

4. Query traffic burst detection instances.
   - Query all traffic burst detection instances.

```
huawei(config)#display traffic-burst-detect instance all

-------------------------------------------------------------------------------
  ID Direction F/S/P          ONT ONT_PORT Period Threshold    state
                                    (ms)   (Kbps)
-------------------------------------------------------------------------------
  11 inbound   0/9/0          - -          10     1000     running
-------------------------------------------------------------------------------

Note: The detection instance based on service-port.
      The F/S/P contains the service flow ID.
```

   - Query a single traffic burst detection instance.

```
huawei(config)#display traffic-burst-detect instance 11

Instance ID          : 11
Direction            : inbound
Position             : port 0/9/0
Period               : 10 ms
Threshold            : 1000 Kbit/s
State                : stopped
Start time           : 2019-09-15 10:03:10+08:00
End time             : 2019-09-15 10:04:10+08:00
ACL match information:
 Matches: ACL 2000 rule 5
```

5. Query the traffic burst detection record.

```
huawei(config)#display traffic-burst-detect result instance 11
It will take some time, please wait...

Instance ID          : 11
Direction            : inbound
Position             : port 0/9/0
Period               : 10 ms
Threshold            : 1000 Kbit/s
State                : stopped
Start time           : 2019-09-15 10:03:10+08:00
End time             : 2019-09-15 10:04:10+08:00
ACL match information:
 Matches: ACL 2000 rule 5
-------------------------------------------------------------------------------
Total traffic burst records: 2
-------------------------------------------------------------------------------
Time                              Traffic BandWidth(Kbit/s)
-------------------------------------------------------------------------------
2019-09-15 10:03:41.0  +08:00                          -
2019-09-15 10:03:40.0  +08:00                          -
```

6. Delete the traffic burst detection instance.

```
huawei(config)#undo traffic-burst-detect instance 11
```

 Copyright © Huawei Technologies Co., Ltd.

13 Multicast

Multicast is a communication mode in which data is transmitted to multiple recipients at the same time.

13.1 Introduction to Multicast

13.2 Basic Multicast Concepts

13.3 Multicast Model

13.4 Implementation Principles of Multicast

13.5 IPv6 Multicast

13.6 Configuring the Multicast Service
The MA5800 supports multicast cascading for reducing the number of ports used on the convergence device, and also supports Multiple Spanning Tree Protocol (MSTP) network protection. With these two functions, the network structure is optimized and the multicast service reliability is improved.

13.7 Multicast Maintenance and Diagnosis

13.8 Multicast Standards and Protocols Compliance

13.9 Multicast Feature Updates

13.1 Introduction to Multicast

Why Is Multicast Required

Traditional IP communications are implemented in unicast or broadcast mode.

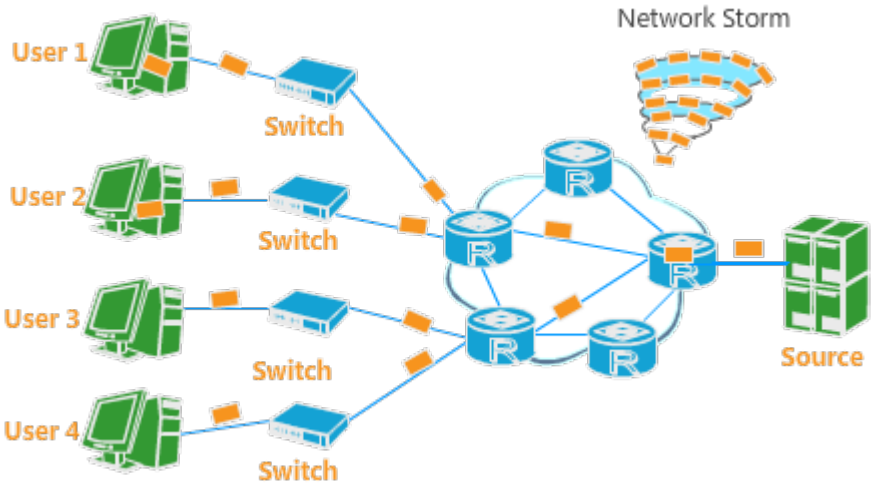
- Unicast: a P2P transmission mechanism. Unicast involves only one information sender and one information recipient.
- Broadcast: a point-to-all-point transmission mechanism. Broadcast involves only one information sender and all the reachable information recipients in a LAN.

In broadcast IP communications, the source sends only one copy of packets to the broadcast address of the network. Then, the network copies these packets and sends them to all network segments for all routers and users, regardless of

 Copyright © Huawei Technologies Co., Ltd.

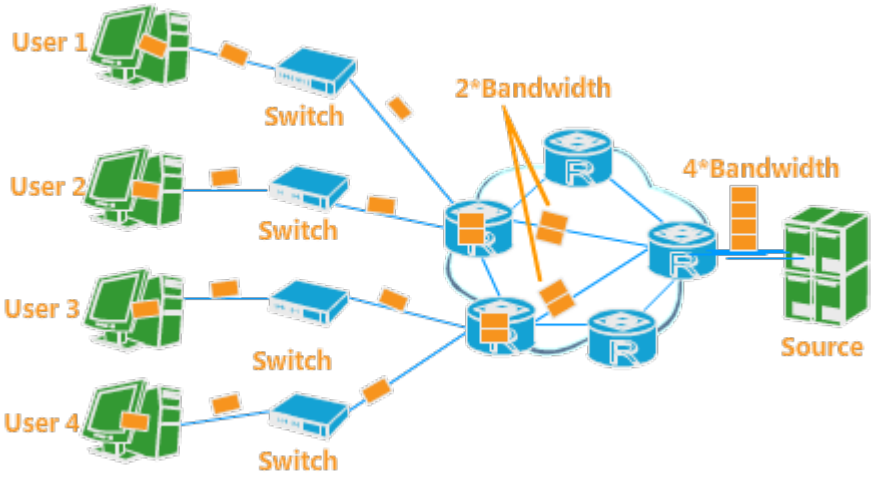
whether these routers and users require such data. In broadcast mode, if only a few users require such packets, network utilization is low and bandwidths are unnecessarily used. The users who do not require such data are adversely affected. In addition, a severe broadcast storm may occur due to routing loopback.

Figure 13-1 Diagram of broadcast implementation



In unicast IP communications, the source sends a separate copy of packets to each recipient. Therefore, each recipient requires a separate data channel. In unicast mode, the amount of data that needs to be transmitted on the network is determined based on the number of users requiring such data. If the number of users requiring the same data is large, multiple same data flows must be set up on the network. In this case, network bandwidth may become insufficient, degrading network transmission quality. Therefore, the unicast mode cannot be used for transmitting large amount of data for a large number of users.

Figure 13-2 Diagram of unicast implementation



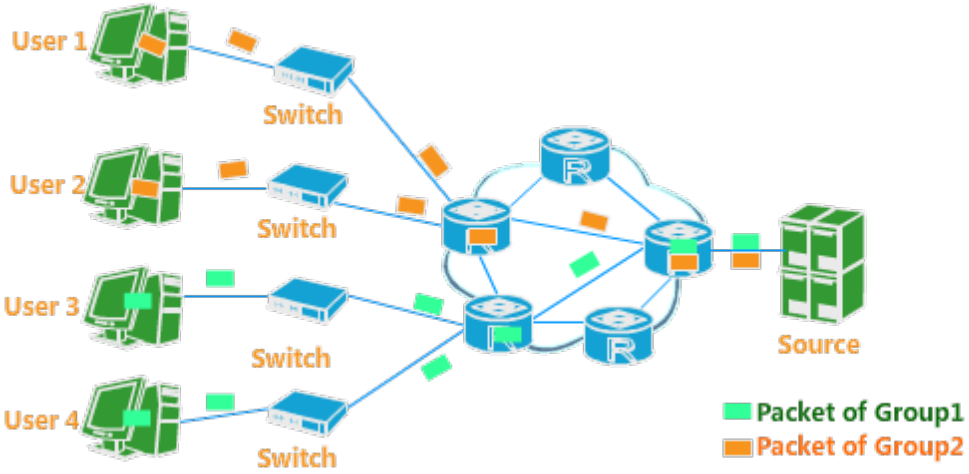
Traditional unicast and broadcast communications cannot meet P2MP service requirements. The IP network development supports more network applications. In addition, some large amount of data needs to be sent in P2MP mode. In this case, multicast is introduced to meet these service requirements.

 Copyright © Huawei Technologies Co., Ltd.

What Is Multicast

Multicast is a communication mode in which one information sender sends one copy of IP data package to a group of information recipients on an IP network. Other hosts on the network cannot receive this data package.

Figure 13-3 Diagram of multicast implementation



Comparison between multicast and unicast, and multicast and broadcast:

- In the multicast mode, a single data stream is sent to a group of users at the same time. Only one copy of the same multicast data stream exists on each link. Compared with the unicast mode, in the multicast mode, the increase of users does not immediately increase the load of the network. Therefore, the server and the CPU can deal with a lighter load, reducing desired network bandwidths.
- Multicast messages can be sent across different network segments and will not be received by users who are not interested in the messages. Compared with the broadcast mode, the multicast mode achieves a longer information transmission distance and ensures that information is transmitted to only interested recipients. Hence, information security can be guaranteed.

The preceding comparisons show that multicast effectively resolves the problem of P2MP transmission and implements efficient P2MP data transmission in IP networks.

Multicast Applications

The multicast technology enables the device to provide value-added services, including live broadcast, IPTV, distance learning, TeleMedicine, network radio, live radio conference, and online game.

13.2 Basic Multicast Concepts

Basic Concepts

The following section provides an example of watching the program of a TV channel to aid the understanding of relevant concepts of IP multicast.

 Copyright © Huawei Technologies Co., Ltd.

- The multicast group is an agreement between the sender and the recipient. For example, a TV channel can be regarded as a multicast group.
- The TV station is the multicast source and it sends data to a certain TV channel.
- The STB is a receiving host. When the user chooses to watch the program of a channel, this action can be regarded as the host joining a multicast group. Then, the TV set displays the program of the TV channel to the user, which means that the host has received the data sent to this multicast group.
- The user can turn on or turn off the STB or switch between channels any time, which means that the host can join or leave a multicast group dynamically.

Figure 13-4 Multicast concept diagram

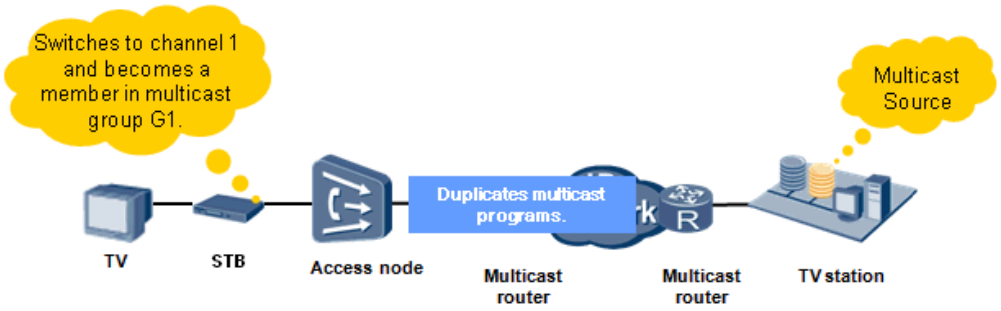


Table 13-1 Table of Basic Concepts

Concept	Description
Multicast group	A multicast group is identified by a multicast IP address. Any host (or any other receiving device) joining a multicast group becomes a member of the group. The group member can identify and receive the IP messages destined to the multicast IP address.
Multicast source	A signal source sending IP messages destined to a multicast address is called a multicast source. • A multicast source can send data to multiple multicast groups at the same time. • Multiple multicast sources can send data to a multicast group at the same time.
Multicast group member	The members of a multicast group are dynamic. Hosts in a network can join or leave a multicast group any time. Multicast group members may be widely dispersed across the network. A multicast source is usually not a data recipient at the same time and is not the member of a multicast group.

 Copyright © Huawei Technologies Co., Ltd.

Concept	Description
Multicast duplication	Multicast duplication is a capability with which the network device duplicates a multicast message from an ingress port into multiple copies and sends them to multiple egress ports. To ensure effective transmission of multitudes of data, this function can be implemented only by hardware.

 Copyright © Huawei Technologies Co., Ltd.

Concept	Description
Multicast address	To enable the communication between a multicast source and its members, a network-layer multicast address must be available, which is the multicast IP address. In addition, a technology must also be available for mapping the multicast IP address to a link-layer multicast MAC address. The following part of this section will describe the two types of multicast address. • Multicast IP address As specified by Internet Assigned Numbers Authority (IANA), multicast messages use class-D IP addresses (224.0.0.0-239.255.255.255) as their destination addresses, and the class-D IP addresses must not appear in the source IP address field of the IP messages. For details of the MAC address range and description, see **Table 13-2**. Multicast addresses are not allocated to the receiving device or the multicast source device for identifying their network location. In the case of the multicast source device, the allocated multicast address is used for generating and carrying multicast data; in the case of the receiving device, the multicast address is used for distinguishing multicast data. In an actual multicast application, the multicast address does not need to be manually input. For example, in the live TV service, which is a common application, a menu interface is provided. When the user orders a program using a remote controller, the application software will automatically obtain the multicast IP address corresponding to the program. • Ethernet multicast MAC address When IP messages are unicast over an Ethernet, the destination MAC addresses used are the MAC addresses of recipients. However, in the transmission of multicast messages, the transmission destination is no longer a specific recipient. Instead, it is a group with uncertain members. In this case, the multicast MAC address is used. Specified by IANA, the most significant 25 bits of a multicast MAC address are 0x01005e, and the least significant 23 bits of the MAC address are the least significant 23 bits of the multicast IP address. The following figure shows the mapping.

 Copyright © Huawei Technologies Co., Ltd.

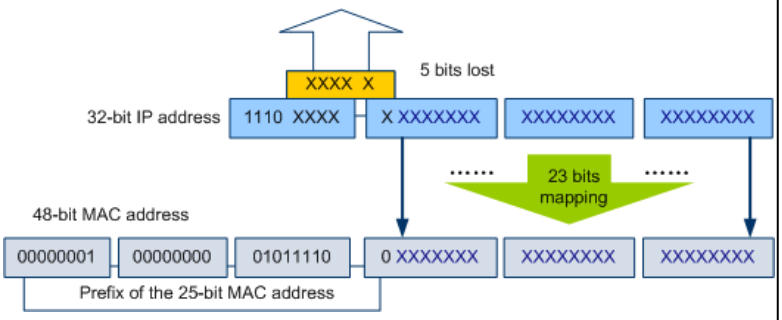
Concept	Description
	Figure 13-5 Mapping between multicast MAC address and multicast IP address  The first four bits of the multicast IP address are 1110, which stands for the multicast ID, and in the last 28 bits, only 23 bits are mapped to the MAC address. Therefore, five bits of information in the IP address is lost. The direct result is that 32 multicast IP addresses are mapped to the same MAC address.
Multicast router	Multicast routers support multicast. A multicast router supports the following functions: • Manages group members in the network segment close to user hosts. • Supports multicast routing for forwarding multicast packets. • Functions as a group member.

Table 13-2 Multicast addresses and meanings

Multicast Address Range	Description
224.0.0.0-224.0.0.255	Permanent group addresses reserved for routing protocols
224.0.1.0-238.255.255.255	User multicast addresses
239.0.0.0-239.255.255.255	Local management group addresses (private addresses) **NOTE** This address range limits multicast to be implemented within a specified multicast domain so that the MAC addresses in different domains can be duplicated.

 Copyright © Huawei Technologies Co., Ltd.

Basic Managed Objects

Table 13-3 Basic managed objects

Managed Object	Description
Multicast VLAN	A multicast VLAN (also called an MVLAN) refers to the VLAN tag carried by multicast data. MVLANs are usually divided based on ISP. By the implementation of the forwarding plane, control plane, and management plane based on VLAN instance, multicast services are provisioned to the users of the same device, allowing the users not to be interfered by each other. Except the super VLAN, the VLAN of any attribute or any type configured on the device can serve as an MVLAN. For details on the MVLAN, see "**13.4.5.6 Multi-instance Multicast**".
Multicast program	A multicast program can be regarded as a multicast group. Its basic attribute is the multicast IP address. The device can manage a multicast program at a finer grain, such as by rights control and CAC. According to whether the attributes (such as the multicast IP address) of each program are configured before the service is provisioned, multicast programs can be classified into two types: pre-configured programs and dynamic programs. Determine whether to support the programs of the two types based on the configured MVLAN program mapping mode. **Table 13-4** shows the supporting status for pre-configured programs and dynamic programs. For details about dynamic programs, see **Dynamic Programs**.
Multicast uplink port	A multicast uplink port is one through which a multicast source connects to the device, and is also a port through which an upper-layer multicast router connects to the device. According to their dependency on the link-layer loop protocol, multicast uplink ports can be classified into two types: manually configured (static) uplink ports and dynamic uplink ports. For details on dynamic uplink ports, see **13.4.4.2 Ring Network of Uplink Ports**.
Multicast user	A multicast user is a multicast data recipient. A service stream must be configured for the multicast user for carrying multicast control messages in the upstream direction (the device can distinguish the user by traffic classification). Therefore, a multicast user corresponds to a unique terminal or service subscriber. In addition, an MVLAN must be specified for the multicast user to indicate to which ISP the service subscriber belongs.

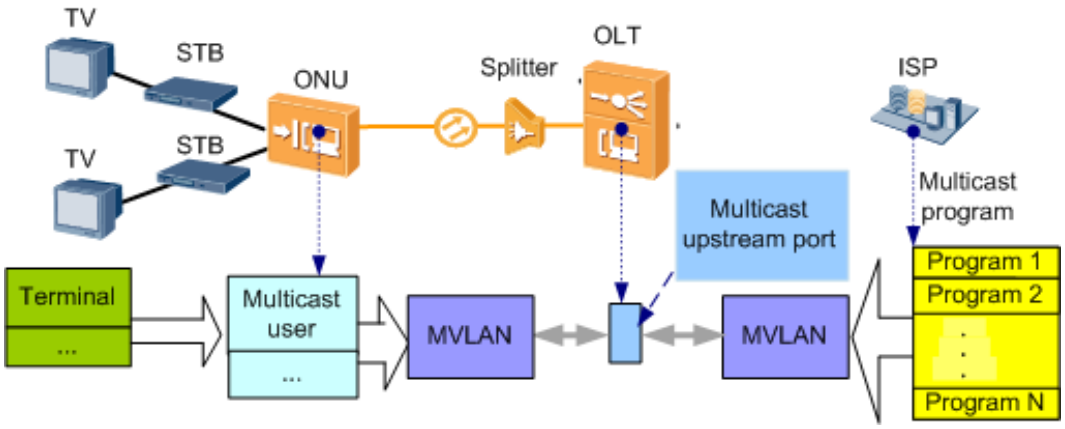
 Copyright © Huawei Technologies Co., Ltd.

Table 13-4 Program type

Program Type	Description
Pre-configured programs	Configure the program list before users watch the video programs. In this mode, the right profile can be used to implement controllable multicast. The program list and the right profile, however, need to be maintained according to the change of the video service. The program host, program prejoin, and multicast bandwidth management functions are supported.
Dynamic programs	A program list is dynamically generated according to the programs requested by users. In this mode, the program list does not need to be configured or maintained; however, the functions such as program management, user multicast bandwidth management, program preview, and program prejoin are not supported.

The following figure shows the relationships between the basic managed objects.

Figure 13-6 Multicast managed objects



NOTE

- As a device placed at users' home, an ONT does not support multiple multicast users. Although the ONT is for only one multicast user, it can still connect to multiple STBs.
- Traffic streams with the QinQ attribute do not support multicast users.
- Traffic streams that classify traffic by double VLANs do not support multicast users.

13.3 Multicast Model

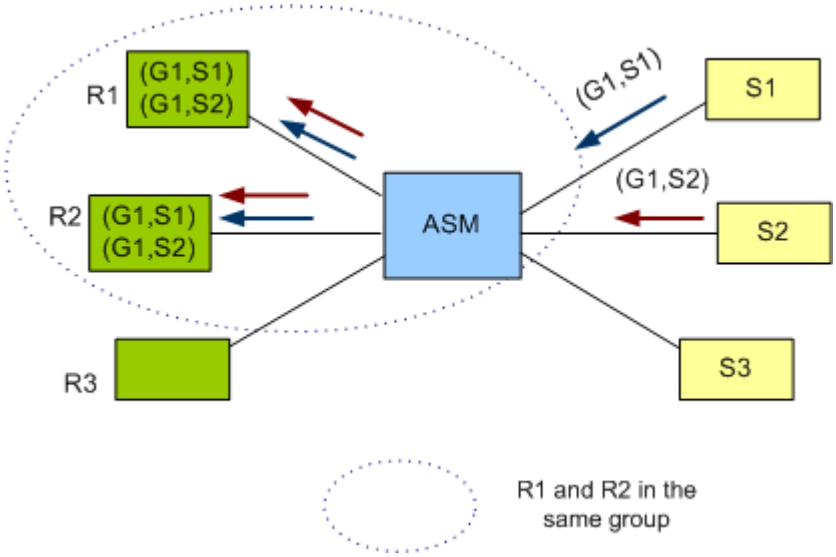
According to the multicast source control level, multicast has three models:

 Copyright © Huawei Technologies Co., Ltd.

ASM

Any-source multicast (ASM) is defined in RFC 1112. In this model, a recipient, by joining a group identified by the multicast address, can receive data sent to the group. A recipient can join or leave a group at any time, and the recipient location or quantity is not limited. In addition, any sender can serve as the multicast source to send data to the group. Therefore, this model is applicable to the multipoint-to-multipoint (MP2MP) multicast application.

Figure 13-7 ASM network model



SFM

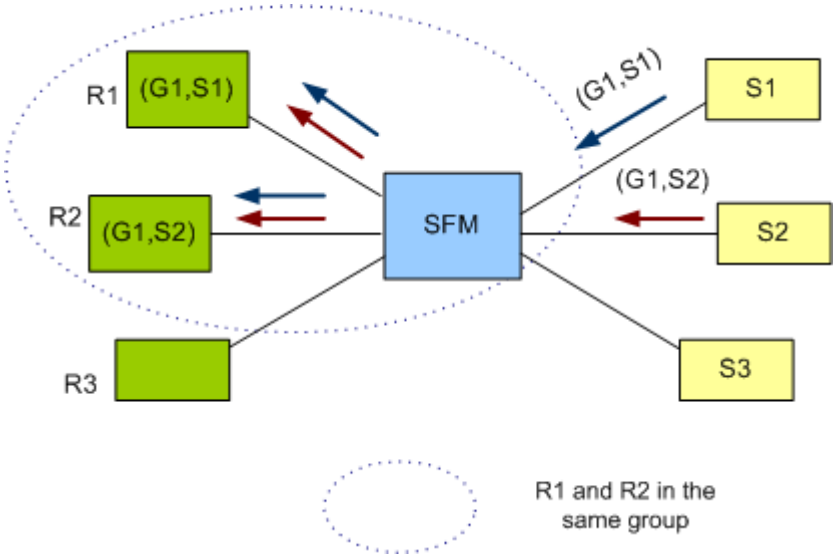
As an extension of ASM, source-filtered multicast (SFM) extends the source filtering function of the upper-layer protocol module. That is, in the SFM model, whether the multicast data of specified multicast source(s) is allowed to pass can be controlled. Viewed from recipients, SFM and ASM are different; but viewed from senders, they are the same. Therefore, SFM is the same as ASM in terms of network interoperability.

NOTE

The SFM is basically an ASM with a multicast source filtering policy. The basic principles and configurations of ASM and SFM are the same. In this manual, both SFM and ASM are called ASM.

 Copyright © Huawei Technologies Co., Ltd.

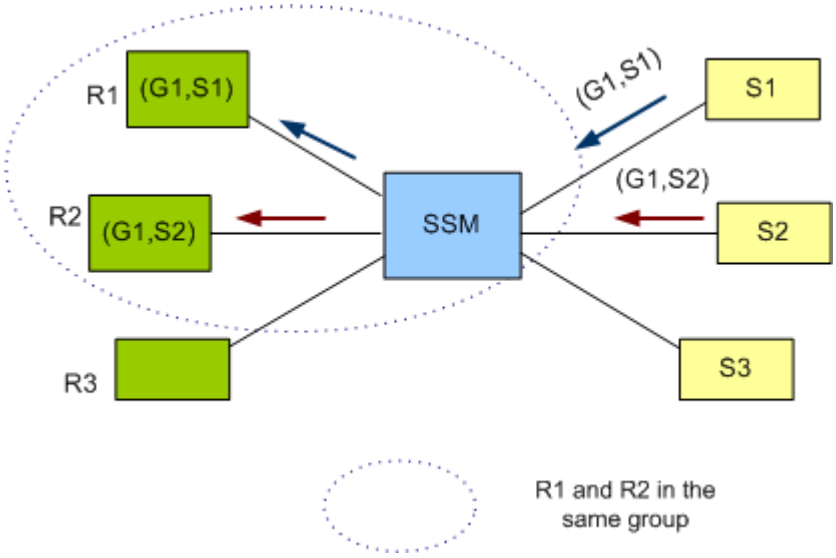
Figure 13-8 SFM network model



SSM

Source-specific multicast (SSM) is defined in RFC 4607. In this model, a recipient joins a channel by specifying the multicast source and group and receives data sent to the group from a specific multicast source. The recipient quantity is not limited. In addition, only the specific sender can serve as the multicast source to send data to the channel. Therefore, this model is applicable to the point-to-multipoint (P2MP) multicast application.

Figure 13-9 SSM network model



The following table lists the protocols that support ASM/SSM.

 Copyright © Huawei Technologies Co., Ltd.

Multicast Model	Typical Protocol Combination of Devices in the Network			
	STB	AN	Router	Inter-domain router
ASM	IGMPv2	IGMPv2	PIM-SM	MSDP/MBGP
SSM	IGMPv3	IGMPv3	PIM-SSM	MBGP

Based on the preceding multicast models, the OLT supports three group filtering modes: ASMSSM, ASM ONLY, and SSM ONLY.

NOTE

Only the OLT supports group filtering mode.

In different group filter modes, for the differences in management plane and control plane, see **Multi-instance Multicast**; for the differences in the forwarding plane, see **Forwarding Framework on the Device**.

Follow the rules below to select a proper filter mode:

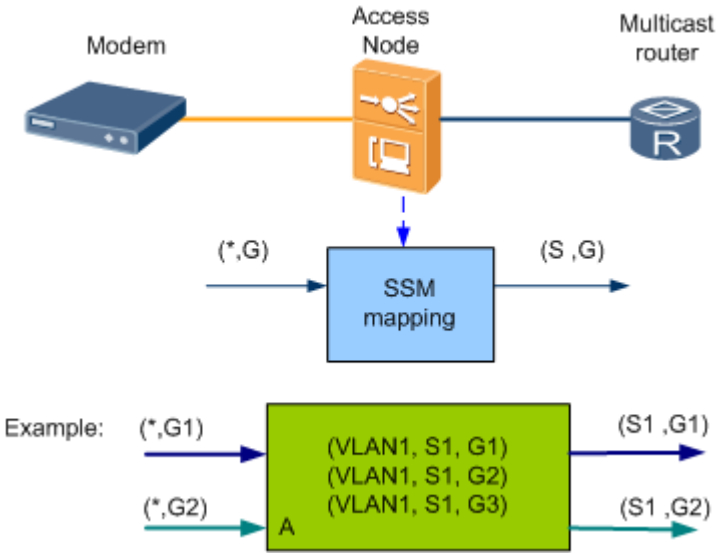
- For compatibility with the original VLAN+GMAC, select ASMSSM.
- For strict ASM or SSM message filtering, select ASM ONLY or SSM ONLY.
- For maximum compatibility of various STBs in a home network, select ASMSSM.
- For SIP+GIP forwarding (that is, posing no restrictions on SIP and GIP planning among different ISPs), select SSM ONLY. However, the entire network needs to support IGMPv3 for implementing SSM ONLY.

When the group filter mode is ASMSSM, even if the user side does not support IGMPv3, carriers can implement SSM network on the network side. The device, with its SSM mapping function (can be supported automatically), can help implement the SSM networking on the network side even if the user-side device does not support IGMPv3. With the SSM mapping function, the device maps the received (*, G) message to an (S, G) message according to the unique multicast program triplet, as shown in the following figure.

- A multicast user cannot belong to multiple MVLANs at the same time.
- Dynamic programs do not support SSM mapping.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-10 SSM mapping



The following table describes whether the devices support the SSM and ASM modes.

User Side	Network Side	Supported or Not
SSM	SSM	Supported
ASM	ASM	Supported
ASM	SSM	Supported
SSM	ASM	Not supported

When the group filter mode is ASM ONLY, SSM mapping is not supported because ASM packets are dropped on the user side. For SSM ONLY, it is unnecessary to support SSM mapping.

13.4 Implementation Principles of Multicast

13.4.1 IGMP

13.4.1.1 IGMP Protocol

Internet Group Management Protocol (IGMP) is used to set up and maintain relationships between multicast members in a multicast group between hosts and the router connected to these hosts.

- Multicast members dynamically add to or leave a multicast group on the host.

 Copyright © Huawei Technologies Co., Ltd.

- The relationships between group members are maintained and managed on the router. In addition, the IGMP protocol supports data exchanging with the upper-layer multicast routing protocol.

IGMP involves three versions: IGMPv1, IGMPv2, and IGMPv3. A new IGMP version is compatible with an earlier version. IGMPv1 is rarely supported. Therefore, according to TR101 requirements, Huawei hosts do not support IGMPv1 and simply discard IGMPv1 packets.

IGMPv1

IGMPv1 is defined in RFC 1112, defining the basic process of group member query and report. IGMPv1 is seldom used. Therefore, IGMPv1 is not described in the remainder of this document.

IGMPv2

IGMPv2 is defined in RFC 2236, supporting quick leave of group members compared with IGMPv1.

Table 13-5 IGMP packet

Role	Message Type	Description
Router	General query	A router periodically sends this message to maintain the requirements posed by all hosts connected to the router in all multicast groups. The router detects an accidentally offline host by an aging mechanism.
	Group-specific query	A router sends this message to check whether a multicast group is still required by any host. The router usually sends this message when receiving a leave message.
Host	Report	The report message is used by a host for actively joining a multicast group or for responding to a general query or a group-specific query.
	Leave	The leave message is used by a host for actively informing a router that the host no longer needs a multicast group.

Group Member Relationship Maintenance

Prerequisite: If a network segment contains multiple multicast routers, IGMPv2 enables the routers to use a querier election mechanism to elect a querier.

1. The querier periodically sends general query messages to group members to query group member relationships.
2. The hosts in the network segment make different responses to the querier after receiving the general query messages.

 Copyright © Huawei Technologies Co., Ltd.

  - The hosts out of the multicast group do not respond to the querier.
  - Group members locally start a timer.
    - If a group member does not monitor a report message responded by other group members when the timer times out, this group member sends a report packet to respond to the querier.
    - If a group member monitors a report message responded by other group members before the timer times out, this group member does not send a report packet to respond to the querier (suppresses its response packet).
3. After receiving the report messages, the querier determines that this network segment contains group members and generates a multicast forwarding entry. When receiving data for the multicast group, the router forwards the data to the group members.

Group Member Join-in

1. If host A added to multicast group G1 monitors no data for G1 in the network segment, host A immediately sends a report message of G1 to the querier, without waiting for the receiving of a general query message sent by the querier.
2. After receiving the report message, the router in the network segment determines that this network segment contains G1 group members and adds the downlink port for G1 to multicast routing entry (*,G1). When receiving data for G1, the router forwards the data to the group members.

Group Member Leaving

1. When host A exits from G1, it sends a leave message to the querier.
2. After receiving the leave message, the querier sends a group-specific query message to group members to check whether all group members in G1 have left this group.
   - If a group member sends a report message to the querier, this network segment still contains G1 group members. Therefore, the querier does not delete the downlink port from multicast entry (*,G1).
   - If no group member sends a report message to the querier, the querier deletes the downlink port from multicast entry (*,G1).

IGMPv3

IGMPv3 is defined in RFC 3376. Compared with IGMPv2 (RFC 2236), IGMPv3 has the following improvements:

- Batch report. The destination IP address of report messages is always filled in as 224.0.0.22. In addition, the IGMP payload can carry multiple group records, reducing the number of report messages between devices. As shown in the following figure, the IGMP message carries the information about two groups 232.1.1.1 and 232.1.1.2 With IGMPv2 messages, the destination IP address must be filled in as the corresponding group IP address. Hence, one IGMPv2 message cannot carry the information about multiple groups.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-11 Example of an IGMPv3 report message

- Longer maximum response time for a query message. In IGMPv3, the maximum response time for the query message is extended from 25.5s (IGMPv2) to 3174.4s. Therefore, IGMPv3 is applicable to large-scale networks.
- Source filtering. With the source filter function, the host can receive or not receive the multicast data carrying the IP address of a specified multicast source. This function enables the device to better implement SSM and support the multiple-ISP scenario. IGMPv2 supports only ASM. The following uses different types of messages to explain the implementation of source filter.
  - Query messages

Message Type	Implementation of Source Filter
General query	The device sends this message to learn the reception status of an interface to "all" multicast groups. This is similar to the general query of IGMPv2.
Group-specific query	The device sends this message to learn the reception status of an interface to the multicast group with a specific address. This is similar to the group-specific query of IGMPv2.
Group-and-source-specific query	The device sends this message to learn the reception status of an interface to the multicast group with a specific group address and source address. This is a new message of IGMPv3.

  - Report messages

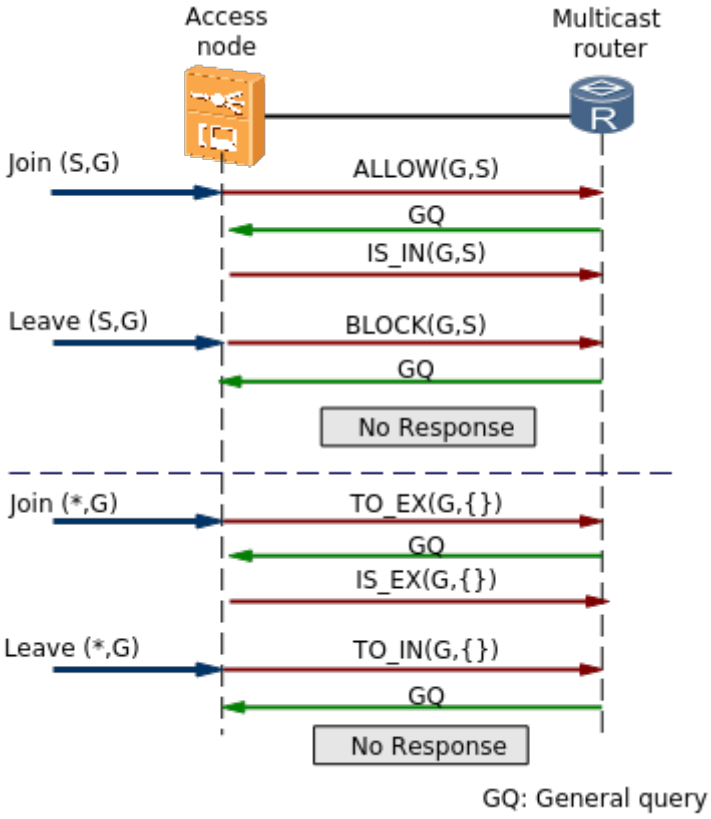
 Copyright © Huawei Technologies Co., Ltd.

Message Type	Implementation of Source Filter
IS_IN(G, S)	Reports the status. Indicates that the current mode of the group is the INCLUDE mode. This message is triggered when the device receives a query message. The source address list contains the source address S of the group.
TO_IN(G, S)	Changes the filter mode of the multicast group to the INCLUDE mode. The source address list contains a new source address S. TO_IN(G, {}) indicates leaving all sources of G and this message in this case is the same as the IGMPv2 leave message.
ALLOW(G, S)	Changes the source address list. This message is triggered when the source address changes. The source address contained in the record is the source address S that the system wishes to join.
BLOCK(G, S)	Changes the source address list. This message is triggered when the source address changes. The source address contained in the record is the source address S that the system does not wish to join.
IS_EX(G, S)	Reports the status. Indicates that the current mode of the group is the EXCLUDE mode. This message is triggered when the device receives a query message. The source address list contains the source address S that the group does not wish to join. IS_EX(G, {}) indicates that the device is interested in all sources of G and this message in this case is the same as the IGMPv2 join message. The device does not support the IS_EX message that contains an empty S.
TO_EX(G, S)	Changes the filter mode of the multicast group to the EXCLUDE mode. The source address list contains a new source address S that the device does not wish to join. TO_EX(G, {}) indicates joining all sources of G and this message in this case is the same as the IGMPv2 join message. The device supports the TO_EX message that contains an empty S.

The following figure shows an example of the report message application.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-12 Program ordering behavior converted into IGMPv3 messages



IGMP Version Compatibility

The compatibility policies of the IGMP version on access devices distinguish between the network side and the user side.

The IGMP version on the network side is configured based on MVLAN. As shown in the following table, according to the IGMP version on the multicast router, the IGMP version on the device should be set to the recommended version to avoid incompatibility. Incompatibility may cause packet loss.

Multicast Router	MVLAN on Access Device	Interoperation Result
v1	v2/v3	Incompatible
v2	v2 (recommended)	Normal
v3	v2 (recommended)	Normal. The device can response the IGMPv3 query messages, and the multicast router can process the IGMPv2 join and leave messages.

 Copyright © Huawei Technologies Co., Ltd.

Multicast Router	MVLAN on Access Device	Interoperation Result
v2	v3	The multicast router does not process IGMPv3 messages. Interoperation is normal only after the IGMP version on the device is downgraded to v2. Before the downgrade, packet loss may occur.
v3	v3 (recommended)	Normal

The IGMP version on the user side can be configured based on multicast users. As shown in the following table, according to the IGMP version on the terminal, the IGMP version on the device should be set to the recommended version to avoid incompatibility. Incompatibility may cause packet loss.

Terminal	Multicast User on Access Device	Interoperation Result
v1	v2/v3/v3-forced	Incompatible
v2	v2 (recommended)	Normal
v3	v2	The device does not process IGMPv3 messages. Interoperation is normal only after the IGMP version on the terminal is downgraded to v2 (the terminal can be downgraded by enabling the function of periodically sending query messages to offline users). Before the downgrade, packet loss may occur.
v2	v3 (recommended)	The terminal does not process IGMPv3 messages. Interoperation is normal only after the IGMP version on the device is downgraded to v2. Even after downgraded to IGMPv2, the device can still identify the IGMPv3 messages sent from other terminals. This ensures greater compatibility of the device. In normal application scenarios, the terminal is usually the active initiating party. The IGMP version on the device can be seamlessly downgraded without packet loss.

 Copyright © Huawei Technologies Co., Ltd.

Terminal	Multicast User on Access Device	Interoperation Result
v3	v3 (recommended)	Normal
v2	v3-forced	Incompatible
v3	v3-forced	The device drops IGMPv2 messages without processing them. Therefore, the device will not be downgraded to v2 but stays in v3.

13.4.1.2 IGMP Mode

MVLAN-based IGMP modes include IGMP proxy and IGMP snooping.

IGMP Proxy

IGMP proxy is a mode in which the device in a tree topology does not set up a route to forward multicast messages, but only acts as a proxy for multicast protocol messages. Details are as follows:

- From the perspective of a terminal, the device serves as a multicast router that implements the functions of the router in the IGMP protocol. Specifically, the device consistently functions as an IGMP querier (not supporting querier election for security concerns) on the user-side network. The device receives and terminates the join and leave messages of all multicast users, and duplicates the multicast program to only the interested multicast users according to the maintained group membership table.

Table 13-6 Structure of the group membership table OLT

Group Filter Mode	Index	Online Member
ASMSSM ASM ONLY	VLAN+GIP	Multicast user list (such as multicast user 1 and multicast user 2)
SSM ONLY	VLAN+GIP+SIP	Multicast user list (such as multicast user 1 and multicast user 2)

- From the perspective of a multicast router, the device serves as a multicast group member that implements the functions of the host in the IGMP protocol. According to the changes (addition or deletion) of the record in the group membership table, the device sends the join message or leave message of a program to the upper layer through the multicast uplink port. In addition, the device responds to the queries of the multicast router according to the status of the group membership table.

 Copyright © Huawei Technologies Co., Ltd.

IGMP proxy effectively reduces the quantity of IGMP messages exchanged on the network side and therefore lessens the load of multicast routers. It is configurable on the device whether to send the IGMP general query to all multicast users or to only interested multicast users.

NOTE

After a multicast user orders a program for the first time, the system generates a value, indicating the random interval. The value ranges from 0 to the difference between the interval for the general query and maximum response time to the general query. After the random interval, the system sends the first general query message to the multicast user.

IGMP Snooping

IGMP snooping enables the device to listen to the multicast protocol packets transmitted between the router and hosts and set up a Layer 2 forwarding table for multicast data packets. This manages and controls multicast data packet forwarding so that the multicast data packets can be sent only to specified recipients. IGMP snooping features:

- Low bandwidth requirements, facilitating separate host charging.
- Separate data forwarding for each VLAN, improving data security.

IGMP snooping has two types:

- IGMP transparent snooping

  It is a snooping function without proxy. The device selects the proxy, snooping, or snooping with proxy function based on MVLANs.

  Before enabling IGMP transparent snooping, run the **igmp query-proxy**, **igmp report-proxy**, and **igmp leave-proxy** commands to disable proxy for query, report, and leave packets, respectively.

  IGMP transparent snooping enables the device to process IGMP messages as follows:

  – Query message

  After receiving the general query message and group-specific query message from the multicast uplink port, the device forwards the network-side query message to the user.

  NOTE

  - To ensure that the multicast user responds to the query in a timely manner, the maximum response time configured on the device must be shorter than that configured on the upper-layer multicast router.
  - The network-side IGMP version of the device is not affected by the multicast router.

  – Join/Leave message

  The device transparently transmits all the join/leave message received from the multicast user to the MVLAN.

  NOTE

  The IGMPv3 message may contain multiple group records that match different MVLANs. In this case, the device segments the message and transparent transmits the segmented messages to the corresponding MVLANs.

 Copyright © Huawei Technologies Co., Ltd.

- IGMP snooping with proxy

  In IGMP upstream transmission, IGMP snooping with proxy is the same as IGMP proxy; in IGMP downstream transmission, however, IGMP snooping with proxy does not suppress the query message as IGMP proxy does.

  IGMP snooping with proxy enables the device to process IGMP messages as follows:

  – Query message

  After receiving the query message from the multicast uplink port, the device reconstructs and sends the query message to the user (default mode, query proxy enabled based on MVLANs) or forwards the query message (disabling query proxy based on MVLANs). It also responds to the multicast router's query according to its multicast group membership table.

  NOTE

  - Like IGMP proxy, the network-side IGMP version of the device is affected by the multicast router.
  - Determine whether to enable query proxy. For details, see the "Usage Guidelines" of the **igmp query-proxy** command.

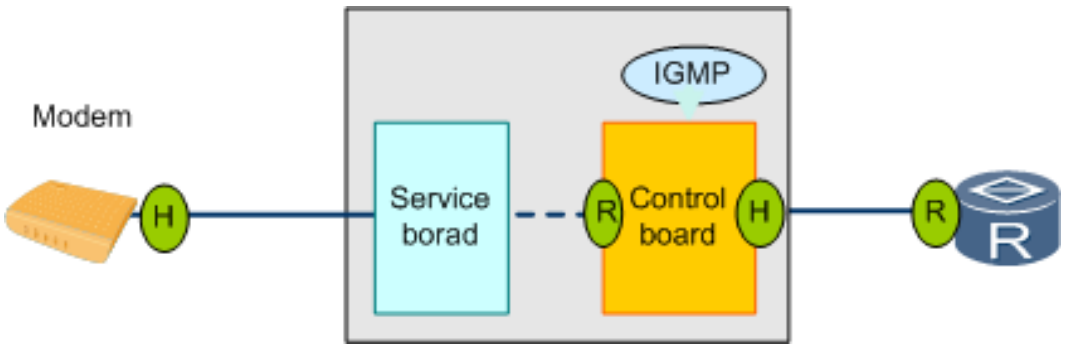
  – Join/Leave message

  The device sends only the first join message from the multicast users to the MVLAN. The device sends only the last leave message from the multicast users to the MVLAN.

13.4.1.3 IGMP Framework on the Device

Architecture of the IGMP Protocol Stack

Figure 13-13 Architecture of the IGMP protocol stack



R represents the router functions of the IGMP protocol, and H represents the host functions of the IGMP protocol.

In the architecture of the IGMP protocol stack, the operation on the control board and the network side is based on MVLAN, and the operation on the user side is based on multicast user (to ensure that the control planes of the users do not interfere with each other).

 Copyright © Huawei Technologies Co., Ltd.

13.4.2 PIM

13.4.2.1 Introduction

Purpose

A multicast network requires multicast protocols to replicate and forward multicast data. PIM is a widely used intra-domain multicast protocol that builds MDTs to transmit multicast data between routers in the same domain.

PIM can create multicast routing entries on demand, forward packets based on multicast routing entries, and dynamically respond to network topology changes.

Definition

PIM is a multicast routing protocol that uses unicast routing protocols to forward data, but PIM is independent of any specific unicast routing protocols.

Concurrently, the MA5800 supports PIM-SM (used in an IPv4 network) and PIM-SSM (used in IPv4 and IPv6 networks).

Table 13-7 PIM implementation modes

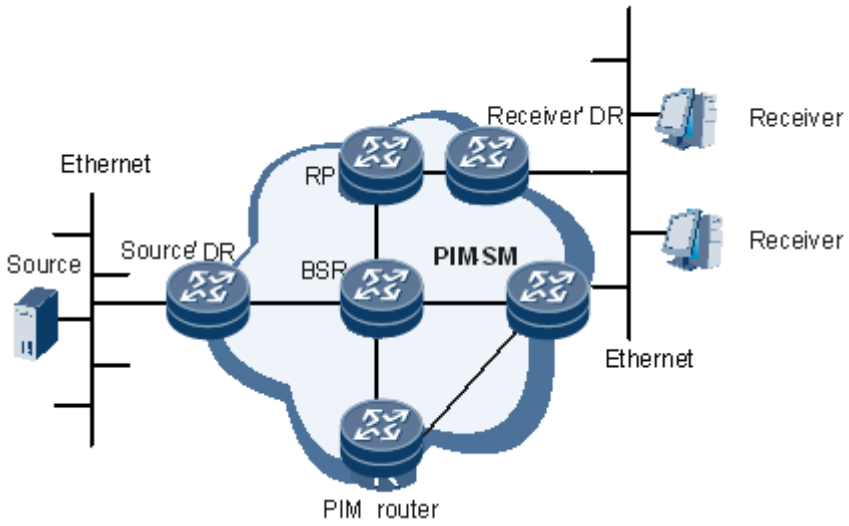
Protocol	Full Name	Model	Deployment Scenario
PIM-SM	Protocol Independent Multicast-Sparse Mode	ASM model	Large-scale networks on which multicast data receivers are sparsely distributed.
PIM-SSM	Protocol Independent Multicast-Source-Specific Multicast	SSM model	Networks on which multicast data receivers can learn source locations before they join multicast groups and require multicast data from specific multicast sources.

Concepts

This section provides basic PIM concepts. **Figure 13-14** shows a typical PIM network.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-14 PIM network



- **PIM device**

  A router that runs PIM is called a PIM device. A router interface on which PIM is enabled is called a PIM interface.

- **Designated router**

  A designated router (DR) can be a multicast source's DR or a receiver's DR.

  – A multicast source's DR is a PIM device directly connected to a multicast source and is responsible for sending Register messages to an RP.

  – A receiver's DR is a PIM device directly connected to receiver's hosts and is responsible for sending Join messages to an RP and forwarding multicast data to receiver's hosts.

  NOTE

  The MA5800 can only be the receiver's DR and cannot be the multicast source's DR.

- **RP**

  An RP is the forwarding core in a PIM-SM domain, used to process hosts' join requests and multicast source's registration requests. An RP constructs an MDT with the RP at the root and creates (S, G) entries to transmit multicast data to hosts. All routers in the PIM-SM domain need to know the RP's location. RP contains static and dynamic RPs.

 Copyright © Huawei Technologies Co., Ltd.

Table 13-8 RP classifications

RP Type	Implementation	Deployment Scenario	Precautions
Static RP	A static RP is manually configured. If a static RP is used, the same RP address must be configured on all PIM devices in the same domain as the RP.	Static RPs are recommended on small-/medium-sized networks because a small-/medium-sized network is stable and has low forwarding requirements for an RP.	To use a static RP, ensure that all PIM devices, including the RP, have the same RP and multicast group address range information.
Dynamic RP	A dynamic RP is elected among C-RPs in the same PIM domain. The BSR sends Bootstrap messages to collect all C-RP information as an RP-Set, and advertises the RP-Set information to all PIM devices in the domain. Then, all the PIM devices use the same RP-Set information and follow the same rules to elect an RP. After the elected RP fails, the other C-RPs start an election process again to elect a new RP.	Dynamic RPs can be used on large-scale networks to improve network reliability and maintainability. • If multiple multicast sources are densely distributed on the network, configuring core devices close to the multicast sources as C-RPs is recommended. • If multiple users are densely distributed on the network, configuring core devices close to the users as C-RPs is recommended.	To use a dynamic RP, you must configure a BSR that dynamically advertises group-to-RP mapping information.

NOTE

The MA5800 supports only static RPs and does not support dynamic RPs or BSR.

- **BSR**

  A BSR on a PIM-SM network collects RP information, summarizes that information into an RP-Set (group-RP mapping database), and advertises the RP-Set to the entire PIM-SM network.

  A network can have only one BSR but can have multiple C-BSRs. If a BSR fails, a new BSR is elected from the C-BSRs.

- **RPT**

  An RPT is an MDT with an RP at the root and group members at the leaves.

 Copyright © Huawei Technologies Co., Ltd.

- **SPT**

  An SPT is an MDT with the multicast source at the root and group members at the leaves.

13.4.2.2 PIM-SM

PIM-SM implements P2MP data transmission on large-scale networks on which multicast data receivers are sparsely distributed. PIM-SM forwards multicast data only to network segments with active receivers that have required the data.

PIM-SM assumes that no host wants to receive multicast data, so PIM-SM sets up an MDT only after a host requests multicast data and then sends the data to the host along the MDT.

Concepts

For details about PIM-SSM, see **Concepts**.

Implementation

The multicast data forwarding process in a PIM-SM domain is as follows:

1. Neighbor discovery

   Each PIM device in a PIM-SM domain periodically sends Hello messages to all other PIM devices in the domain to discover PIM neighbors and maintain PIM neighbor relationships.

   By default, a PIM device permits other PIM control messages or multicast messages from a neighbor, irrespective of whether the PIM device has received Hello messages from the neighbor. However, if a PIM device has the neighbor check function, the PIM device permits other PIM control messages or multicast messages from a neighbor only after the PIM device has received Hello messages from the neighbor.

2. DR election

   PIM devices exchange Hello messages to elect a DR on a shared network segment. The receiver's DR is the only multicast data forwarder on a shared network segment. The source's DR is responsible for forwarding multicast data received from the multicast source along an MDT.

   NOTE

   The MA5800 does not support DR election. It supports DR by default.

3. RP discovery

   An RP is the forwarding core in a PIM-SM domain. An RP forwards multicast data over the entire network.

   NOTE

   The MA5800 supports only static RPs. It is specified by configuring the same RP address on all PIM devices in a network and therefore it does not need election.

4. RPT setup

   PIM-SM assumes that no hosts want to receive multicast data, so PIM-SM sets up an RPT only after a host requests multicast data, and then sends the data from the RP to the host along the RPT.

 Copyright © Huawei Technologies Co., Ltd.

5. SPT switchover

   A multicast group in a PIM-SM domain is associated with only one RP and one RPT. All multicast data packets are forwarded by the RP. The path along which the RP forwards multicast data may not be the shortest path from the multicast source to receivers. The load of the RP increases when the multicast traffic volume increases. If the multicast data forwarding rate exceeds a configured threshold, an RPT-to-SPT switchover can be implemented to reduce the burden on the RP.

   NOTE

   The MA5800 supports packet forwarding upon receiving the first packet (equivalent to threshold 0), and does not support threshold-crossing switchover (switchover is performed when the traffic reaches an extent).

If a network problem occurs, the Assert mechanism can be used to guarantee that multicast data is transmitted properly. If multiple multicast data forwarders exist on a network segment, each multicast packet is repeatedly sent across the network segment, generating redundant multicast data. To resolve this issue, the Assert mechanism can be used to select a unique multicast data forwarder on a network segment.

The detailed PIM-SM implementation process is as follows:

Neighbor Discovery

Each PIM-enabled interface on a PIM device sends Hello messages. A multicast packet that carries a Hello messages has the following features:

- The destination address is 224.0.0.13, indicating that this packet is destined for all PIM devices on the same network segment as the interface that sends this packet.
- The source address is an interface address.
- The TTL value is 1, indicating that the packet is sent only to neighbor interfaces.

Hello messages are used to discover neighbors, adjust protocol parameters, and maintain neighbor relationships.

- Discovering PIM neighbors

  All PIM devices on the same network segment must receive multicast packets with the destination address 224.0.0.13. Directly connected multicast devices can then learn neighbor information from the received Hello messages.
- Adjusting protocol parameters

  A Hello message carries the following protocol parameters:
  - DR_Priority: priority used by each device to elect a DR. The higher a device's priority is, the higher the probability that the device will be elected as the DR.
  - Holdtime: timeout period during which the neighbor remains in the reachable state.
  - LAN_Delay: delay for transmitting a Prune message on the shared network segment.
  - Override-Interval: interval carried in a Hello message for overriding a Prune message.

 Copyright © Huawei Technologies Co., Ltd.

- Maintaining neighbor relationships

  PIM devices periodically exchange Hello messages. If a PIM device does not receive a new Hello messages from its PIM neighbor within the Holdtime, the device considers the neighbor unreachable and deletes the neighbor from its neighbor list.

  PIM neighbor relationship changes cause the multicast topology to change. If an upstream or a downstream neighbor is unreachable, multicast routes re-converge, and the MDT is transferred.

RPT Setup

Figure 13-15 RPT setup and data forwarding processes

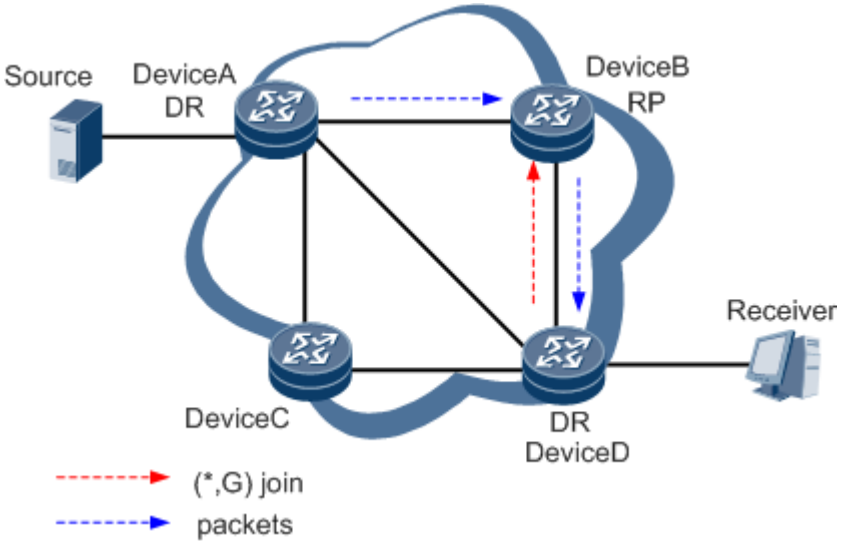


Figure 13-15 shows the RPT setup and data forwarding processes.

- When a multicast source sends the first multicast packet of a multicast group to its DR, the source's DR encapsulates the multicast packet in a Register message and unicasts the Register message to the RP. The RP creates an (S, G) entry to register the multicast source information.
- When a receiver joins a multicast group through IGMP, the receiver's DR sends a Join message to the RP. A (*, G) entry is then created on each hop, and an RPT is created.
- When a receiver joins a multicast group and a multicast source becomes active simultaneously, the multicast source's DR encapsulates the multicast packet in a Register message and unicasts the Register message to the RP. The RP then forwards the multicast data along the RPT to group members.

The RPT implements on-demand multicast data forwarding, which reduces bandwidth usage by preventing unrequested data transmission.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

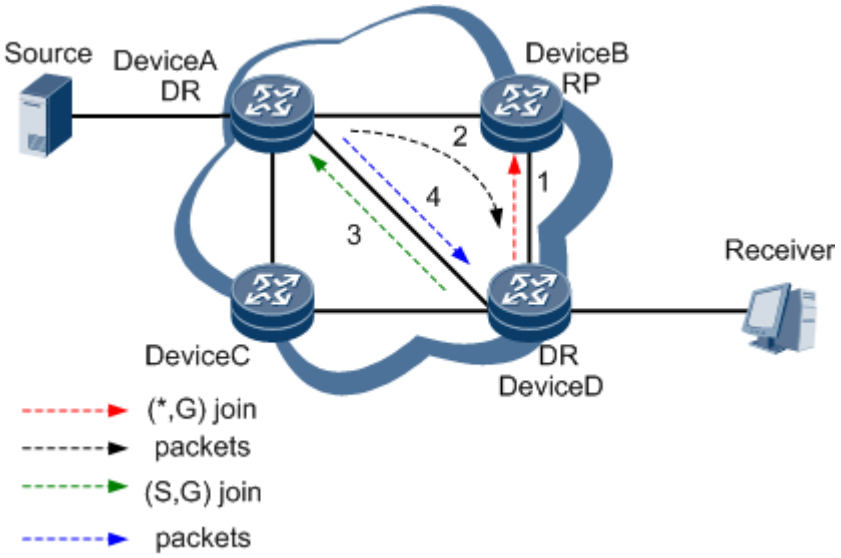
To reduce the RPT forwarding loads and improve multicast data forwarding efficiency, PIM-SM supports SPT switchover, allowing a multicast network to set up an SPT with the multicast source at the root. Then, the multicast source can send multicast data directly to receivers along the SPT.

SPT Switchover

In a PIM SM domain, a multicast group interacts with only one RP, and only one RPT is set up. If SPT switchover is not enabled, all multicast packets must be encapsulated in Register messages and then sent to the RP. After receiving the packets, the RP de-encapsulates them and forwards them along the RPT.

Since all multicast packets forwarded along the RPT are transferred by the RP, the RP may be overloaded when multicast traffic is heavy. To resolve this problem, PIM-SM allows the RP or the receiver's DR to trigger an SPT switchover.

Figure 13-16 SPT switchover triggered by the receiver's DR



An SPT switchover can be triggered by the RP or by the receiver's DR:

- SPT switchover triggered by the RP

  Register messages sent from the source's DR are decapsulated by the RP, which then forwards multicast data along the RPT to group members. At the same time, the RP sends SPT Join messages to the source's DR to set up an SPT from the RP to the source.

  After the SPT is set up and starts carrying multicast data packets, the RP stops processing Register messages. This frees the source's DR and RP from encapsulating and decapsulating packets. Multicast data is sent from the device directly connected to the multicast source to the RP along the SPT and then forwarded to group members along the RPT.
- SPT switchover triggered by the receiver's DR

 Copyright © Huawei Technologies Co., Ltd.

a. As shown in **Figure 13-16**, multicast data is transmitted along the RPT. The receiver's DR (Device D) sends (*, G) Join messages to the RP. Multicast data is sent to receivers along the path source's DR (Device A) -> RP (Device B) -> receiver's DR (Device D).

b. The receiver's DR periodically checks the forwarding rate of multicast packets. If the receiver's DR finds that the forwarding rate is greater than the configured threshold, the DR triggers an SPT switchover.

c. The receiver's DR sends (S, G) Join messages to the source's DR. After receiving multicast data along the SPT, the receiver's DR discards multicast data received along the RPT and sends a Prune message to the RP to delete the receiver from the RPT. The switchover from the RPT to the SPT is complete.

d. Multicast data is forwarded along the SPT. Specifically, multicast data is transmitted to receivers along the path multicast source's DR (Device A) -> receiver's DR (Device D).

An SPT is set up from the source to group members, and therefore subsequent packets may bypass the RP. The RPT may not be an SPT. After an SPT switchover is performed, delays in transmitting multicast data on the network are reduced.

If one source sends packets to multiple groups simultaneously and an SPT switchover policy is specified for a specified group range:

- Before an SPT switchover, these packets reach the receiver's DR along the RPT.
- After an SPT switchover, only the packets sent to the groups within the range specified in the SPT switchover policy are forwarded along the SPT. Packets sent to other groups are still forwarded along the RPT.

NOTE

By default, the RP performs an SPT switchover immediately after receiving the first Register message, and the receiver's DR performs an SPT switchover immediately after receiving the first multicast packet.

Assert

Either of the following conditions indicates other multicast forwarders are present on the network segment:

- A multicast packet fails the RPF check.
- The interface that receives the multicast packet is a downstream interface in the (S, G) entry on the local device.

If other multicast forwarders are present on the network segment, the PIM device starts the Assert mechanism.

The device sends an Assert message through the downstream interface. The downstream interface also receives an Assert message from a different multicast forwarder on the network segment. The destination address of the multicast packet in which the Assert message is encapsulated is 224.0.0.13. The source address of the packet is the downstream interface address. The TTL value of the packet is 1. The Assert message carries the route cost from the PIM device to the source or RP, priority of the used unicast routing protocol, and the group address.

 Copyright © Huawei Technologies Co., Ltd.

The PIM device compares its information with the information contained in the message sent by its neighbor. This is called Assert election. The election rules are as follows:

1. The device that runs a higher priority unicast routing protocol wins.
2. If the devices have the same unicast routing protocol priority, the device with the smaller route cost to the source wins.
3. If the devices have the same priority and route cost, the device with the highest IP address for the downstream interface wins.

The device performs the following operations based on the Assert election result:

- If the device wins the election, the downstream interface of the device is responsible for forwarding multicast packets on the network segment. The downstream interface is called an Assert winner.
- If the device does not win the election, the downstream interface is prohibited from forwarding multicast packets and is deleted from the downstream interface list of the (S, G) entry. The downstream interface is called an Assert loser.

After Assert election is complete, only one upstream device that has a downstream interface exists on the network segment, and the downstream interface transmits only one copy of each multicast packet. The Assert winner then periodically sends Assert messages to maintain its status as the Assert winner. If the Assert loser does not receive any Assert messages from the Assert winner after the timer of the Assert loser expires, the loser re-adds downstream interfaces for multicast data forwarding.

13.4.2.3 PIM-SSM

Protocol Independent Multicast-Source-Specific Multicast (PIM-SSM) enables a user host to rapidly join a multicast group if the user knows a multicast source address. PIM-SSM sets up a shortest path tree (SPT) from a multicast source to a multicast group, while PIM-SM uses rendezvous points (RPs) to set rendezvous point trees (RPTs). Therefore, PIM-SSM implements a more rapid join function than PIM-SM.

Different from the any-source multicast (ASM) model, the SSM model does not need to maintain an RP, construct an RPT, or register a multicast source.

The SSM model is based on PIM-SM and IGMPv3. The procedure for setting up a multicast forwarding tree on a PIM-SSM network is similar to the procedure for setting up an SPT on a PIM-SM network. The receiver's DR, which knows the multicast source address, sends Join messages directly to the source so that multicast data streams can be sent to the receiver's designated router (DR).

Related Concepts

For details about PIM-SSM, see **Concepts**.

Implementation

The process for forwarding multicast data in a PIM-SSM domain is as follows:

 Copyright © Huawei Technologies Co., Ltd.

1. Neighbor Discovery

   Each PIM device in a PIM-SSM domain periodically sends Hello messages to all other PIM devices in the domain to discover PIM neighbors and maintain PIM neighbor relationships.

   By default, a PIM device permits other PIM control messages or multicast messages from a neighbor, irrespective of whether the PIM device has received Hello messages from the neighbor. However, if a PIM device has the neighbor check function, the PIM device permits other PIM control messages or multicast messages from a neighbor only after the PIM device has received Hello messages from the neighbor.

2. DR Election

   PIM devices exchange Hello messages to elect a DR on a shared network segment. The receiver's DR is the only multicast data forwarder on the segment.

   NOTE

   The MA5800 does not support DR election. It supports DR by default.

3. SPT setup

   Users on a PIM-SSM network can know the multicast source address and can, therefore, specify the source when joining a multicast group. After receiving a Report message from a user, the receiver's DR sends a Join message towards the multicast source to establish an SPT between the source and the user. Multicast data is then sent by the multicast source to the user along the SPT.

   SPT establishment can be triggered by user join requests (both dynamic and static) and SSM-mapping.

   The DR in an SSM scenario is valid only in the shared network segment connected to group members. The DR on the group member side sends Join messages to the multicast source, creates the (S, G) entry hop by hop, and then sets up an SPT.

If a network problem occurs, the Assert mechanism can be used to guarantee that multicast data is transmitted properly. If multiple multicast data forwarders exist on a network segment, each multicast packet is repeatedly sent across the network segment, generating redundant multicast data. To resolve this issue, the Assert mechanism can be used to select a unique multicast data forwarder on a network segment.

The detailed PIM-SSM implementation process is as follows:

Neighbor Discovery

Each PIM-enabled interface on a PIM device sends Hello messages. A multicast packet that carries a Hello messages has the following features:

- The destination address is 224.0.0.13, indicating that this packet is destined for all PIM devices on the same network segment as the interface that sends this packet.
- The source address is an interface address.
- The TTL value is 1, indicating that the packet is sent only to neighbor interfaces.

 Copyright © Huawei Technologies Co., Ltd.

Hello messages are used to discover neighbors, adjust protocol parameters, and maintain neighbor relationships.

- Discovering PIM neighbors

  All PIM devices on the same network segment must receive multicast packets with the destination address 224.0.0.13. Directly connected multicast devices can then learn neighbor information from the received Hello messages.

- Adjusting protocol parameters

  A Hello message carries the following protocol parameters:

  – DR_Priority: priority used by each device to elect a DR. The higher a device's priority is, the higher the probability that the device will be elected as the DR.

  – Holdtime: timeout period during which the neighbor remains in the reachable state.

  – LAN_Delay: delay for transmitting a Prune message on the shared network segment.

  – Override-Interval: interval carried in a Hello message for overriding a Prune message.

- Maintaining neighbor relationships

  PIM devices periodically exchange Hello messages. If a PIM device does not receive a new Hello messages from its PIM neighbor within the Holdtime, the device considers the neighbor unreachable and deletes the neighbor from its neighbor list.

  PIM neighbor relationship changes cause the multicast topology to change. If an upstream or a downstream neighbor is unreachable, multicast routes re-converge, and the MDT is transferred.

Assert

Either of the following conditions indicates other multicast forwarders are present on the network segment:

- A multicast packet fails the RPF check.
- The interface that receives the multicast packet is a downstream interface in the (S, G) entry on the local device.

If other multicast forwarders are present on the network segment, the PIM device starts the Assert mechanism.

The device sends an Assert message through the downstream interface. The downstream interface also receives an Assert message from a different multicast forwarder on the network segment. The destination address of the multicast packet in which the Assert message is encapsulated is 224.0.0.13. The source address of the packet is the downstream interface address. The TTL value of the packet is 1. The Assert message carries the route cost from the PIM device to the source or RP, priority of the used unicast routing protocol, and the group address.

The PIM device compares its information with the information contained in the message sent by its neighbor. This is called Assert election. The election rules are as follows:

1. The device that runs a higher priority unicast routing protocol wins.

 Copyright © Huawei Technologies Co., Ltd.

2. If the devices have the same unicast routing protocol priority, the device with the smaller route cost to the source wins.
3. If the devices have the same priority and route cost, the device with the highest IP address for the downstream interface wins.

The device performs the following operations based on the Assert election result:

- If the device wins the election, the downstream interface of the device is responsible for forwarding multicast packets on the network segment. The downstream interface is called an Assert winner.
- If the device does not win the election, the downstream interface is prohibited from forwarding multicast packets and is deleted from the downstream interface list of the (S, G) entry. The downstream interface is called an Assert loser.

After Assert election is complete, only one upstream device that has a downstream interface exists on the network segment, and the downstream interface transmits only one copy of each multicast packet. The Assert winner then periodically sends Assert messages to maintain its status as the Assert winner. If the Assert loser does not receive any Assert messages from the Assert winner after the timer of the Assert loser expires, the loser re-adds downstream interfaces for multicast data forwarding.

13.4.2.4 PIM Control Messages

PIM devices exchange control messages to implement multicast routing. A PIM control message is encapsulated in an IP packet, as shown in **Figure 13-17**.

Figure 13-17 Encapsulation format of a PIM control message

- The protocol type field is 103.
- The destination address identifies a receiver. The destination address can be either a unicast address or a multicast address.

PIM Control Message Types

All PIM control messages use the same header format, as shown in **Figure 13-18**.

Figure 13-18 Header format of a PIM protocol message

 Copyright © Huawei Technologies Co., Ltd.

Table 13-9 Fields in a PIM control message

Field	Description
Version	PIM version The value is 2.
Type	Message type: • 0: Hello • 1: Register • 2: Register-Stop • 3: Join/Prune • 5: Assert
Reserved	Reserved
Checksum	Checksum

Hello Messages

PIM devices periodically send Hello messages through all PIM interfaces to discover neighbors and maintain neighbor relationships.

In an IP packet that carries a Hello message, the source address is a local interface's address, the destination address is 224.0.0.13, and the TTL value is 1. The IP packet is transmitted in multicast mode.

Figure 13-19 Hello message format

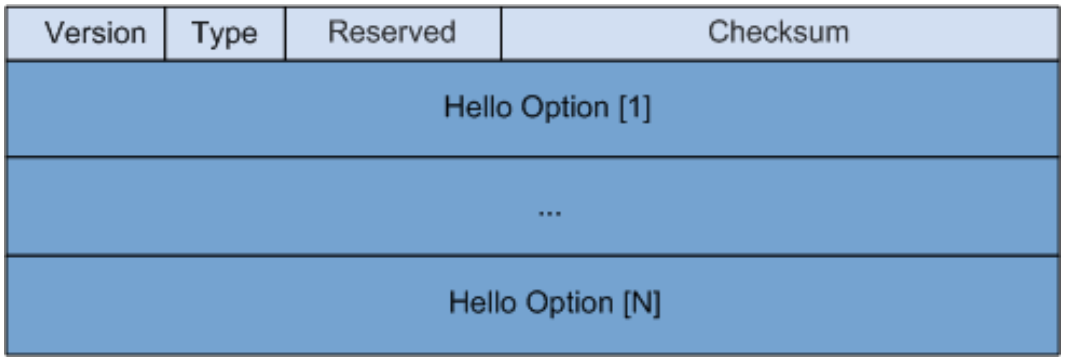
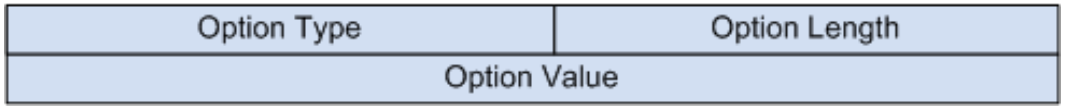


Figure 13-20 Hello Option field format



 Copyright © Huawei Technologies Co., Ltd.

Table 13-10 Fields in a Hello message

Field	Description
Type	Message type The value is 0.
Reserved	Reserved The field is set to 0 when the message is sent and is ignored when the message is received.
Checksum	Checksum
Option Type	Parameter type Valid values of this field are listed in Table 1-13.
Option Length	Length of the Option Value field
Option Value	Parameter value

Table 13-11 Valid values of the Option Type field

Option Type	Option Value
1	Holdtime: timeout period during which a neighbor remains in the reachable state If a device does not receive any Hello message during the timeout period, the device considers its neighbor unreachable.
2	The field consists of the following parts: • LAN Prune Delay: delay before transmitting Prune messages on a shared network segment • Interval: interval for overriding a Prune message • T: capability of suppressing Join messages
19	DR Priority: priority of a device interface, used to elect a designated router (DR) The higher a device interface's priority, the higher the probability the device interface becomes a DR.
20	Generation ID: a random number, indicating neighbor status If the neighbor status changes, the random number is updated. When a device detects that the Hello messages received from an upstream device contain different Generation ID values, the device considers the upstream neighbor Down or the status of the upstream neighbor has changed.

 Copyright © Huawei Technologies Co., Ltd.

Option Type	Option Value
21	State Refresh Capable: interval for refreshing neighbor status
22	Bidirectional Capable PIM-Hello: If a message sent by a device contains this parameter, this device is capable of Bidirectional Protocol Independent Multicast (BIDIR-PIM). The length of the Option Value field is 0.
24	Address List: secondary address list of PIM interfaces

Register Messages

NOTE

Register messages are used only in PIM-SM.

When a multicast source becomes active on a PIM-SM network, the source's DR sends a Register message to register with the rendezvous point (RP).

In an IP packet that carries a Register message, the source address is the address of the source's DR, and the destination address is the RP's address. The message is transmitted in unicast mode.

Figure 13-21 Register message format

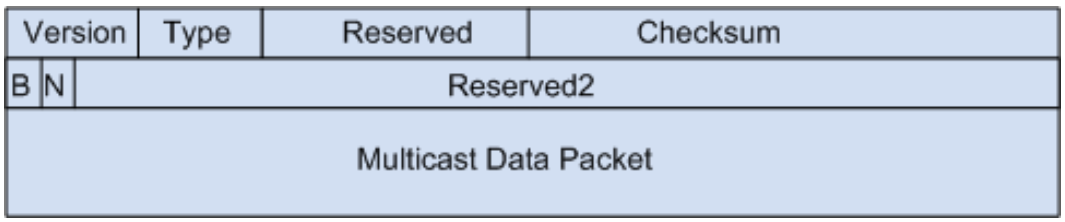


Table 13-12 Fields in a Register message

Field	Description
Type	Message type The value is 1.
Reserved	The field is set to 0 when the message is sent and is ignored when the message is received.
Checksum	Checksum
B	Border bit
N	Null-Register bit

 Copyright © Huawei Technologies Co., Ltd.

Field	Description
Reserved2	Reserved. The field is set to 0 when the message is sent and is ignored when the message is received.
Multicast data packet	The source's DR encapsulates the received multicast data in a Register message and sends the message to the RP. After decapsulating the message, the RP learns the (S, G) information of the multicast data packet.

A multicast source can send data to multiple groups, and therefore a source's DR must send Register messages to the RP of each target multicast group. A Register message is encapsulated only in one multicast data packet, so the packet carries only one copy of (S, G) information.

In the register suppression period, a source's DR sends Null-Register messages to notify the RP of the source's active state. A Null-Register message contains only an IP header, including the source address and group address. After the register suppression times out, the source's DR encapsulates a Register message into a multicast data packet again.

Register-Stop Messages

NOTE

Register-Stop messages are used only in PIM-SM.

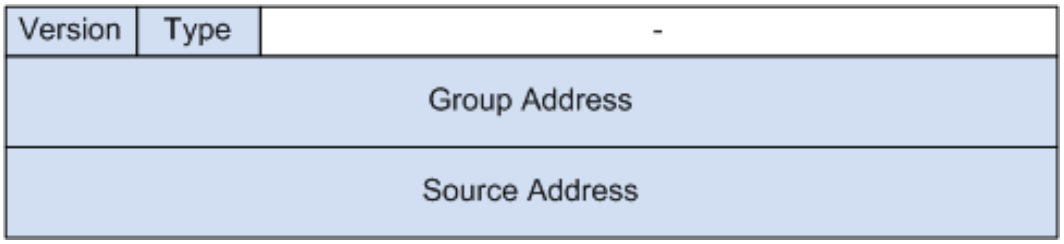
On a PIM-SM network, an RP sends Register-Stop messages to a source's DR in the following conditions:

- Receivers stop requesting a multicast group's data from the RP.
- The RP stops serving a multicast group.
- Multicast data has been switched from a rendezvous point tree (RPT) to a shortest path tree (SPT).

After receiving a Register-Stop message, a source's DR stops using the Register message to encapsulate multicast data packets and enters the register suppressed state.

In an IP packet that carries a Register-Stop message, the source address is the RP's address, and the destination address is the source DR's address. The message is transmitted in unicast mode.

Figure 13-22 Register-Stop message format



 Copyright © Huawei Technologies Co., Ltd.

Table 13-13 Fields in a Register-Stop message

Field	Description
Type	Message type The value is 2.
Group Address	Multicast group address
Source Address	Multicast source address

An RP can serve multiple groups, and a group can receive data from multiple sources. Therefore, an RP may simultaneously perform multiple (S, G) registrations.

A Register-Stop message carries only one copy of the (S, G) information. When an RP sends a Register-Stop message to a source's DR, the RP can terminate only one (S, G) registration.

After receiving the Register-Stop message carrying the (S, G) information, the source's DR stops encapsulating (S, G) packets. The source still uses Register messages to encapsulate packets and send the packets to other groups.

Join/Prune Messages

A Join/Prune message can contain both Join messages and Prune messages. A Join/Prune message that contains only a Join message is called a Join message. A Join/Prune message that contains only a Prune message is called a Prune message.

- When a PIM device is not required to send data to its downstream interfaces, the PIM device sends Prune messages through its upstream interfaces to instruct upstream devices to stop forwarding packets to the network segment on which the PIM device resides.
- When a receiver starts to require data from a PIM-SM network, the receiver's DR sends a Join message through the reverse path forwarding (RPF) interface towards the RP to instruct the upstream neighbor to forward packets to the receiver. The Join message is sent in the upstream direction hop by hop to set up an RPT.
- When an RP triggers an SPT switchover, the RP sends a Join message through the RPF interface connected to the source to instruct the upstream neighbor to forward packets to the network segment. The Join message is sent in the upstream direction hop by hop to set up an SPT.
- When a receiver's DR triggers an SPT switchover, the DR sends a Join message through the RPF interface connected to the source to instruct the upstream neighbor to forward packets to the network segment. The Join message is sent in the upstream direction hop by hop to set up an SPT.
- A PIM network segment may be connected to a downstream interface and multiple upstream interfaces. After an upstream interface sends a Prune message, if other upstream interfaces still require multicast packets, these

 Copyright © Huawei Technologies Co., Ltd.

interfaces must send Join messages within the override-interval. Otherwise, the downstream interfaces responsible for forwarding packets on the network segment do not perform the prune action.

NOTE

- If PIM is enabled on the interfaces of user-side routers, a receiver' DR is elected, and outbound interfaces are added to the PIM DR's outbound interface list. The PIM DR then sends Join messages to the RP.
- If PIM is not enabled on the interfaces of the user-side routers, no receiver' DR is elected, and outbound interfaces are added to the IGMP querier's outbound interface list. The IGMP sends Join messages to the RP.

In an IP packet that carries a Join/Prune message, the source address is a local interface's address, the destination address is 224.0.0.13, and the TTL value is 1. The message is transmitted in multicast mode.

Figure 13-23 Join/Prune messages on a PIM shared network segment

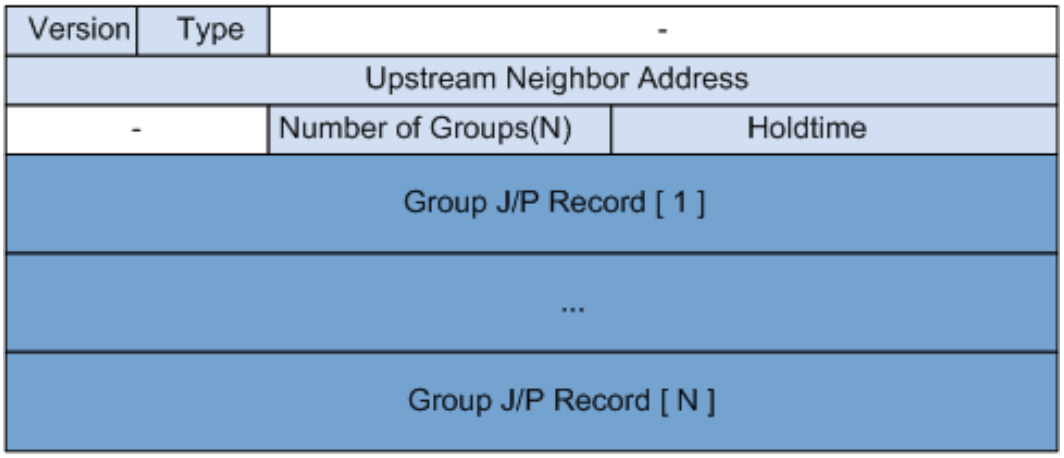
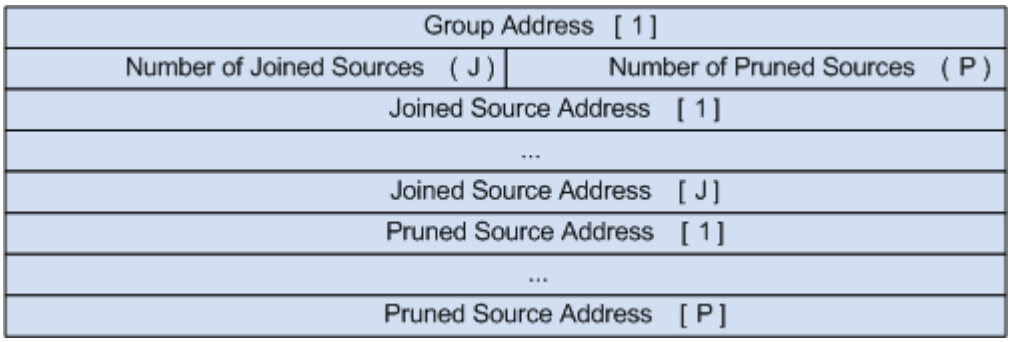


Figure 13-24 Format of the Group J/P Record field



 Copyright © Huawei Technologies Co., Ltd.

Table 13-14 Fields in a Join/Prune message

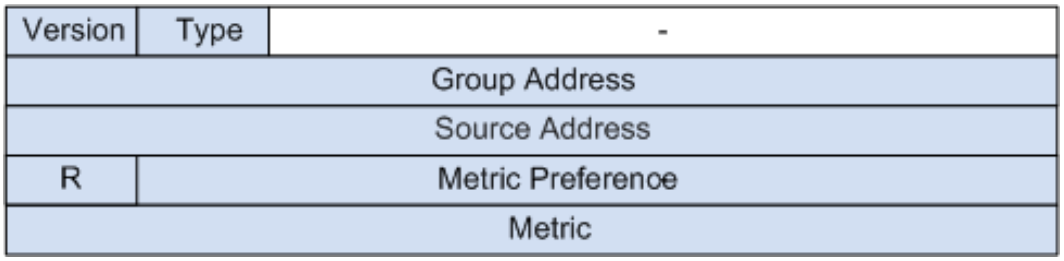
Field	Description
Type	Message type The value is 3.
Upstream Neighbor Address	Upstream neighbor's address, that is, the address of the downstream interface that receives the Join/Prune message and performs the Join or Prune action
Number of Groups	Number of groups contained in the message
Holdtime	Duration (in seconds) that an interface remains in the Join or Prune state
Group Address	Group address
Number of Joined Sources	Number of sources that the device joins
Number of Pruned Sources	Number of sources that the device prunes
Joined Source Address	Address of the source that the device joins
Pruned Source Address	Address of the source that the device prunes

Assert Messages

On a shared network segment, if a PIM device receives an (S, G) packet from the downstream interface of the (S, G) or (*, G) entry, it indicates that other forwarders exist on the network segment. The PIM device then sends an Assert message through the downstream interface to participate in the forwarder election. The device that fails in the forwarder election stops forwarding multicast packets through the downstream interface.

In an IP packet that carries an Assert message, the source address is a local interface's address, the destination address is 224.0.0.13, and the TTL value is 1. The packet is transmitted in multicast mode.

Figure 13-25 Assert message format



 Copyright © Huawei Technologies Co., Ltd.

Table 13-15 Fields in an Assert message

Field	Description
Type	Message type The value is 5.
Group Address	Group address
Source address	This field is a multicast source address if a unique forwarder is elected for (S, G) entries, and this field is 0 if a unique forwarder is elected for (*, G) entries.
R	RPT bit This field is 0 if a unique forwarder is elected for (S, G) entries, and this field is 1 if a unique forwarder is elected for (*, G) entries.
Metric Preference	Priority of the unicast path to the source address If the R field is 1, this field indicates the priority of the unicast path to the RP.
Metric	Cost of the unicast route to the source address If the R field is 1, this field indicates the cost of the unicast path to the RP.

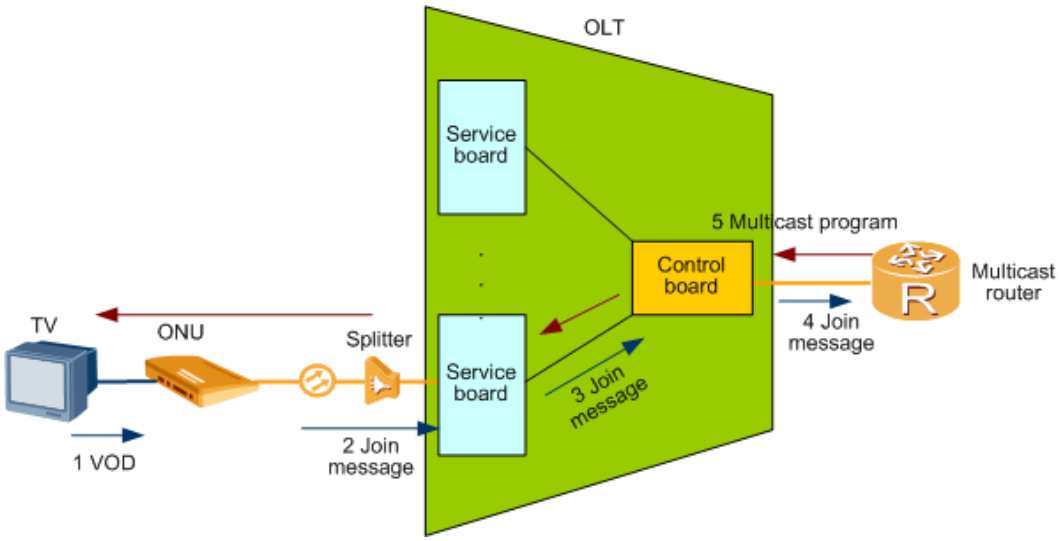
13.4.3 Multicast Forwarding

13.4.3.1 Join Flow

This section considers IGMPv2 proxy as an example to describe the join flow.

Join Flow

Figure 13-26 Join flow



 Copyright © Huawei Technologies Co., Ltd.

1. The multicast user switches a channel and sends a join message for demanding a new program GIP1.
2. After receiving the join message, the control board enters the IGMP protocol stack of the multicast user. After multicast control is implemented (for details, see "**13.4.5.3 Multicast CAC**"), the following group membership table is generated on the control board.

Index	Online Member
MVLAN1+GIP1	Multicast user 1

At the same time, the following multicast forwarding tables are generated on the control board and the service board (for details on how to map GIP1 to GMAC1, see "**Basic Concepts**").

Table 13-16 Multicast forwarding table on the control board

Index	Duplication Destination
MVLAN1+ GMAC1	Corresponding port on service board 1

Table 13-17 Multicast forwarding table on the service board

Index	Duplication Destination
MVLAN1+ GMAC1	User port 1

3. According to MVLAN1 corresponding to the program, the service board serves as the proxy of multicast user 1 and sends a join message to the control board.
4. The control board then sends a join message to the multicast router through the multicast upstream port of MVLAN1.
5. After receiving the multicast stream, the device first duplicates the stream to service board 1 according to the multicast forwarding table of the control board, and then duplicates the stream to user port 1 according to the multicast forwarding table of the service board.

NOTE

Though the SVLAN of a multicast user is different from the MVLAN, the device can still implement the mapping to the MVLAN according to the multicast member configuration relationship. In this way, cross-VLAN multicast is supported without requiring additional configuration.

13.4.3.2 Leave Flow

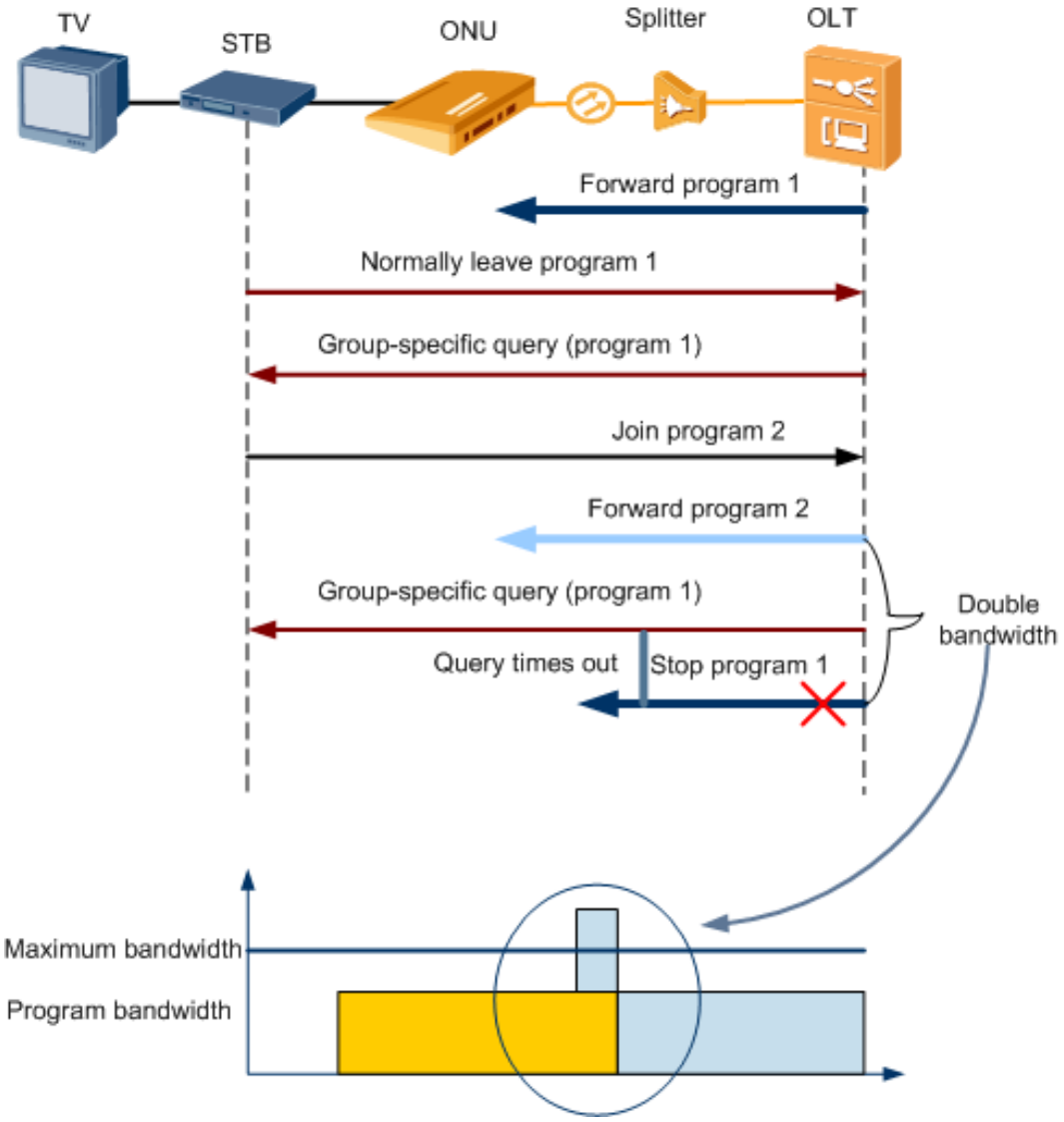
Normal leave

As defined by IGMPv2, the router must send a group-specific query message after it receives the leave message from a host, and it considers that the host does not

 Copyright © Huawei Technologies Co., Ltd.

need the data of the group until the query times out. The following figure illustrates the flow of a normal leave (the same to IGMPv3).

Figure 13-27 Flow of a normal leave



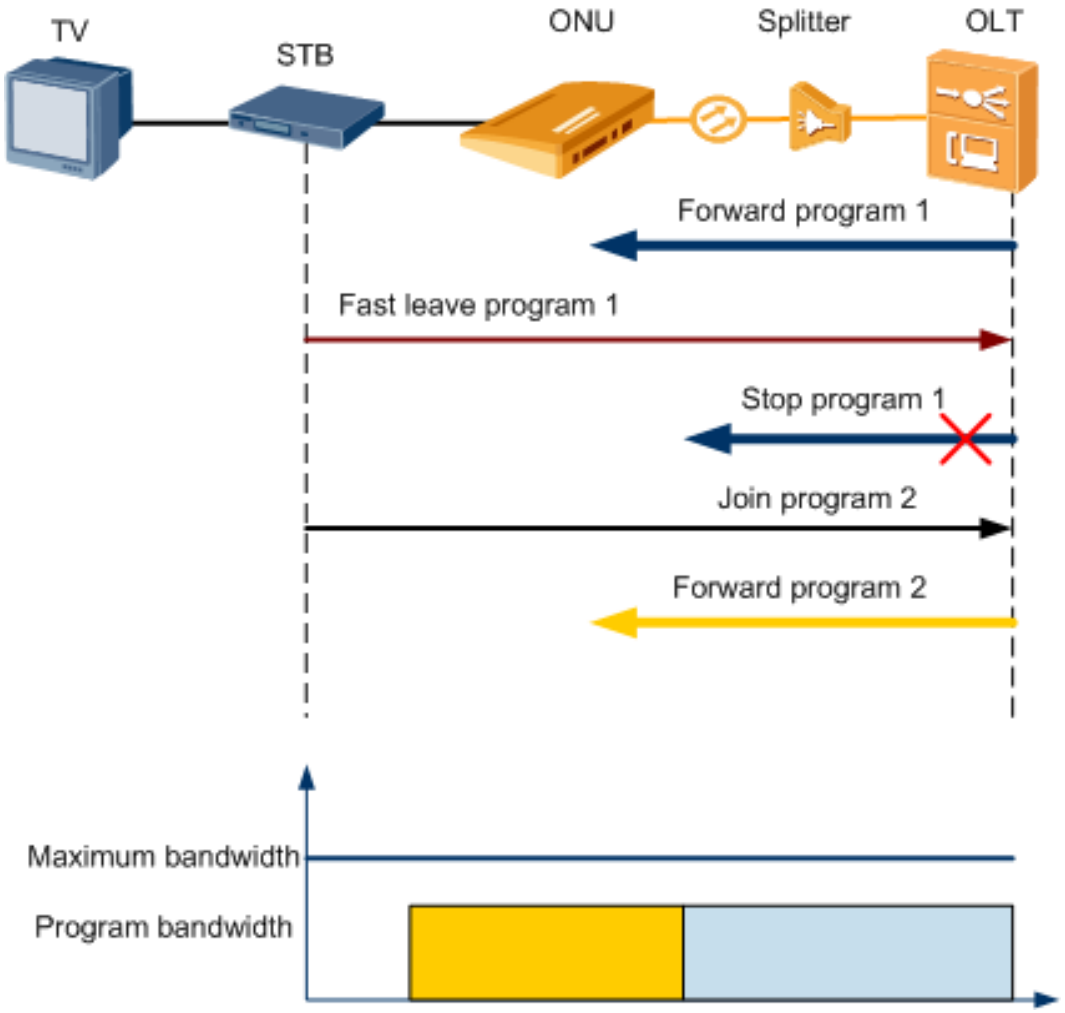
Two IGMP messages are transmitted in the case of a channel switching, one for leaving the original multicast group and one for joining the new multicast group. Therefore, traffic of two multicast groups exists on the subscriber line before the original multicast group is stopped. If the subscriber line does not reserve sufficient bandwidth for carrying the traffic of two multicast groups, traffic overflow (packet loss) will occur. For example, if video streams are carried, pixelation will occur.

Quick leave

When the device receives the leave message from a multicast user, it immediately stops forwarding the messages of the user. The following figure illustrates the flow of a quick leave (the same to IGMPv3).

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-28 Flow of a quick leave

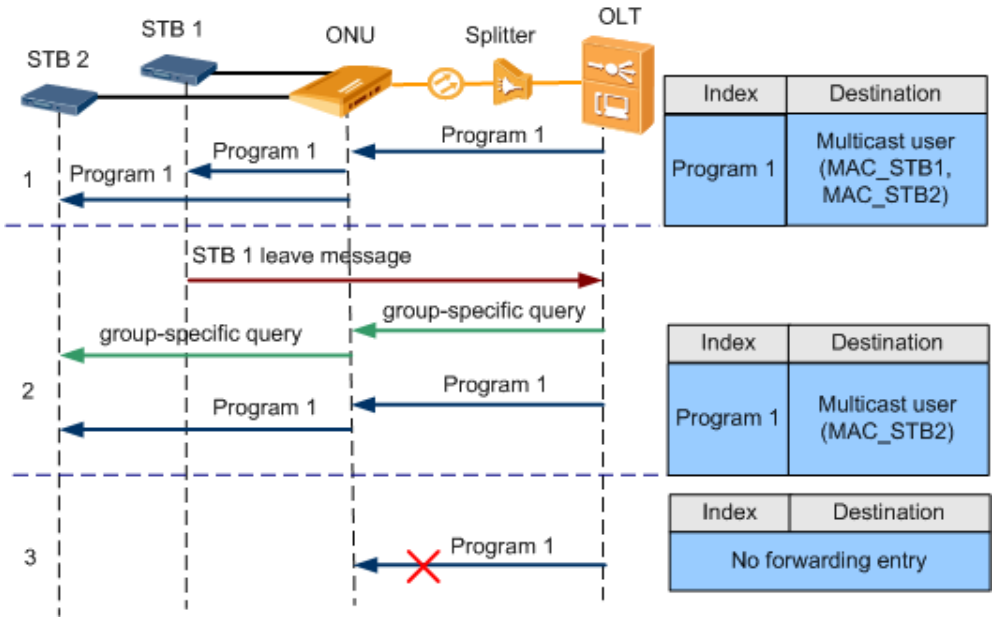


Quick leave based on the MAC address

When the device generates a multicast group membership table, it not only records the multicast user but also records the MAC addresses of the multicast group members of the multicast user. A maximum of eight MAC addresses are supported for each multicast user. When the device receives a leave message, it first deletes the MAC addresses in the multicast group membership table, and it stops forwarding the messages of the group only when all the MAC addresses of the multicast user are deleted. The following figure illustrates the flow of a quick leave based on the MAC address (the same to IGMPv3).

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-29 Flow of a quick leave based on the MAC address (STB2 has silently left.)

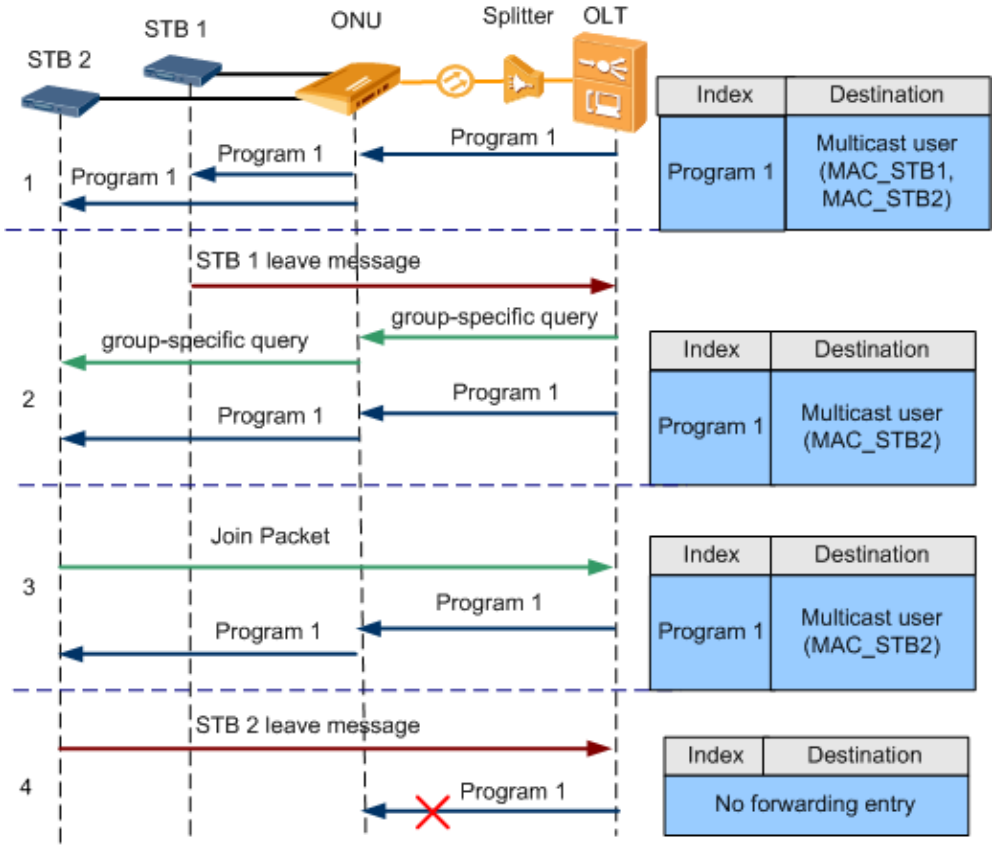


NOTE

The silent leaving of an STB indicates that the STB exceptionally leaves a program. For example, the STB is powered off or goes offline.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-30 Flow of a quick leave based on the MAC address (STB2 is online.)



In summary, the three leave modes have their advantages and disadvantages. You can choose any of them according to actual needs and can set the leave mode for a multicast user.

Leave Mode	User-side Multi-STB Supported or Not	Bandwidth Occupation Time
Normal leave	Yes The STB quantity is not limited.	Aged upon reception of group-specific query
Quick leave	No	Released immediately
Quick leave based on the MAC address	Yes One multicast user supports a maximum of eight STBs at a time.	Released immediately

The following configurations are recommended and can be adopted by different users according to their home network topologies.

 Copyright © Huawei Technologies Co., Ltd.

HG Function	STB Quantity	Reserved Bandwidth	Normal Leave	Quick Leave	Quick Leave Based on the MAC Address
No IGMP	One	Insufficient	-	√	√
		Sufficient	√	√	√
	Several	Insufficient	-	-	√ (Less than eight)
		Sufficient	√	-	√ (Less than eight)
IGMP snooping	One	Insufficient	-	√	√
		Sufficient	√	√	√
	Several	Insufficient	-	-	√ (Less than eight)
		Sufficient	√	-	√ (Less than eight)
IGMP proxy	One	Insufficient	-	√	√ (Not limited)
		Sufficient	√	√	√ (Not limited)
	Several	Insufficient	-	√	√ (Not limited)
		Sufficient	√	√	√ (Not limited)

Global Leave

As defined in TR101, the global leave message is an IGMP message with an all-zero group IP address, which indicates leaving all the groups.

- Network side

When the network topology changes, the device sends the global leave message to the upper-layer multicast router. After receiving the message, the upper-layer multicast router immediately sends the general query message, with the maximum response time set to the maximum time of responding to the group-specific query message. The device, after receiving the query message, responds to the upper-layer multicast router with the join message of the interested group. In this way, the multicast service can recover more quickly. Here, the network

 Copyright © Huawei Technologies Co., Ltd.

topology change events include ring network switching, line up/down, and active/standby port switching in a protect group.

NOTE

- If the device is interconnected with a network device that does not support the global leave message, multicast services may be interrupted during the network topology change. Therefore, it is recommended that the global leave function be manually disabled on the device.
- The device supports sending of the global leave message only in IGMPv2.

- User side

When the STB is powered on immediately after a sudden power-off, because the STB cannot remember the previously-watched program, the bandwidth of the previously-watched program and the program resources are released only after the general query ages.

If the STB supports the global leave function, the STB sends a global leave message after it is re-powered on. After receiving the message, the device sends a general query message, with the maximum response time set to the maximum time of responding to the group-specific query message. If the multicast user is a fast-leave or MAC-based fast-leave user, the device releases all program resources of this multicast user. If the user is a normal-leave user, the device sends a group-specific query message and releases the program resources after the group-specific query times out.

NOTE

Only the IGMPv2 global leave messages can be processed.

13.4.3.3 Forwarding Framework on the Device

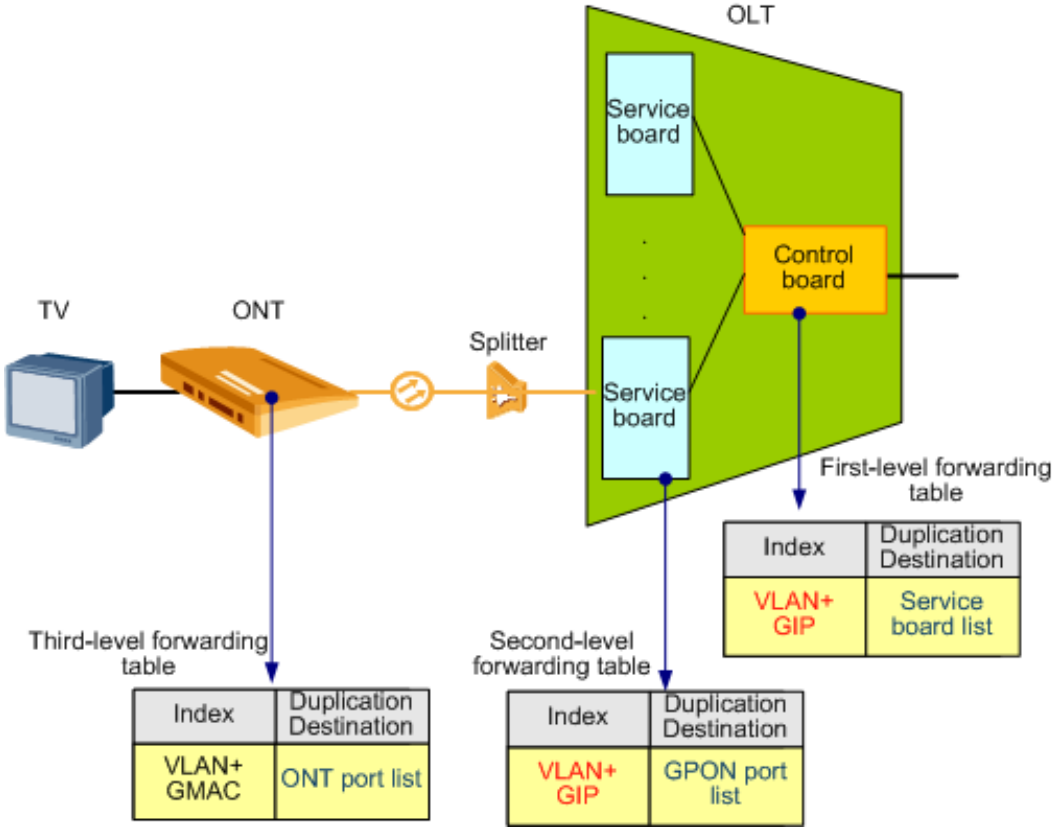
NOTE

The difference among the following forwarding tables is the first-level and second-level indexes. For example, the first-level and second-level index in the GPON multicast forwarding table (VLAN+GIP) is VLAN+GIP.

 Copyright © Huawei Technologies Co., Ltd.

GPON Multicast Forwarding Table (VLAN+GIP) [OLT]

Figure 13-31 GPON multicast forwarding table (VLAN+GIP)



The OLT supports a distributed 2+1-level duplication architecture:

- The first-level duplication is implemented on the control board. By using the "VLAN+GIP" index, the control board duplicates multicast data to the service board interested in the multicast program in an as-per-requirement manner, effectively saving the backplane bandwidth.
- The second-level duplication is implemented on the service board. By using the "VLAN+GIP" index, the service board duplicates multicast data to the GPON port interested in the multicast program in an as-per-requirement manner, effectively saving the downstream bandwidth of the GPON port. Then the service board encapsulates and transmits the multicast data on the GPON port in the mode of multicast GEM port.
- The third-level duplication is implemented on the ONT. By using the "VLAN +GMAC" white list, the ONT filters out unneeded multicast data to avoid bandwidth overflow at the downstream ingress (ONT only supports that in olt-control mode). Then, by using the "VLAN+GMAC" index, the ONT duplicates the multicast data to the ONT ports in an as-per-requirement manner (only supports forwarding by using GMAC in snooping mode).

 Copyright © Huawei Technologies Co., Ltd.

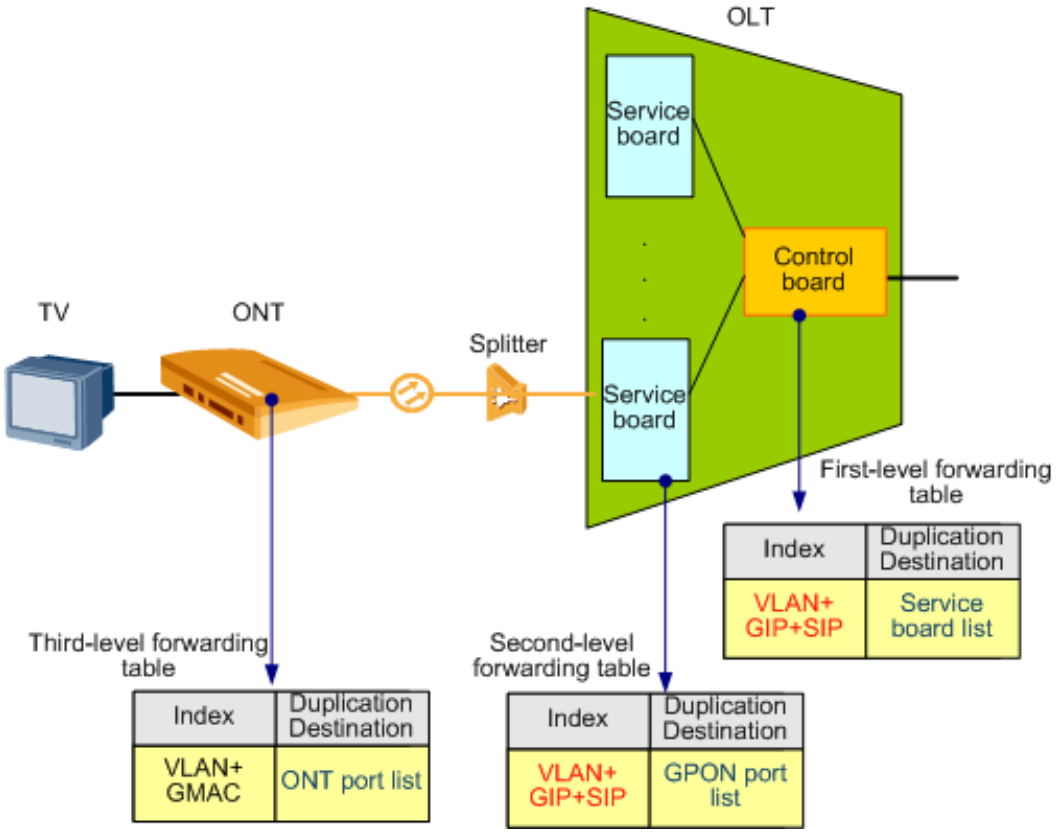
NOTE

1. This forwarding mode applies to ASMSSM and ASM ONLY group filter modes (for details, see "**group filtering mode**").
2. The ONT that implements the third-level duplication is recommended to use the chip that supports VLAN+GIP forwarding. If the ONT does not use such a chip, program GIP-to-GMAC mapping must be a one-to-one mapping. Without the one-to-one mapping on the same ONT, garbled images will occur.

This topic describes only the forwarding framework in the most common single-copy duplication mechanism. For the hardware forwarding framework in the multi-copy duplication mechanism, see "**GPON Multi-Copy Duplication**."

GPON Multicast Forwarding Table (VLAN+GIP+SIP) [OLT]

Figure 13-32 GPON multicast forwarding table (VLAN+GIP+SIP)



The OLT supports a distributed 2+1-level duplication architecture:

- The first-level duplication is implemented on the control board. By using the "VLAN+GIP+SIP" index, the control board duplicates multicast data to the service board interested in the multicast program in an as-per-requirement manner, effectively saving the backplane bandwidth.
- The second-level duplication is implemented on the service board. By using the "VLAN+GIP+SIP" index, the service board duplicates multicast data to the GPON port interested in the multicast program in an as-per-requirement manner, effectively saving the downstream bandwidth of the GPON port. Then the service board encapsulates and transmits the multicast data on the GPON port in the mode of multicast GEM port.

 Copyright © Huawei Technologies Co., Ltd.

- The third-level duplication is implemented on the ONT. By using the "VLAN +GMAC" white list, the ONT filters out unneeded multicast data to avoid bandwidth overflow at the downstream ingress (ONT only supports that in olt-control mode). Then, by using the "VLAN+GMAC" index, the ONT duplicates the multicast data to the ONT ports in an as-per-requirement manner (only supports forwarding by using GMAC in snooping mode).

NOTE

1. This forwarding mode applies to the SSM ONLY group filter mode.
2. Due to hardware limitations, only the difference in the least significant 20 bits can be differentiated between source IP addresses (SIPs). For example, 1.1.1.1 and 2.1.1.1 are the same SIP for the device; 1.1.1.1 and 1.1.1.2 are different SIPs for the device.
3. The ONT that implements the third-level duplication is recommended to use the chip that supports VLAN+GIP forwarding. If the ONT does not use such a chip, program GIP-to-GMAC mapping must be a one-to-one mapping. Without the one-to-one mapping on the same ONT, garbled images will occur.

13.4.3.4 GPON Multicast Duplication

If the service board of the OLT is a GPON service board, the device has two multicast forwarding mechanisms and you can configure the forwarding mechanism based on MVLAN.

- Single-copy duplication

  This is the most common duplication mode of GPON multicast (it also refers to the mode mentioned in this document unless otherwise stated). This mode makes the best of the GPON downstream WDM transmission principle and sends multicast data to all ONTs using the non-encrypted GEM port bearer channel. Each ONT receives the multicast data according to the multicast filtering table. For the corresponding hardware forwarding entry, see "**Multicast Forwarding Table**."

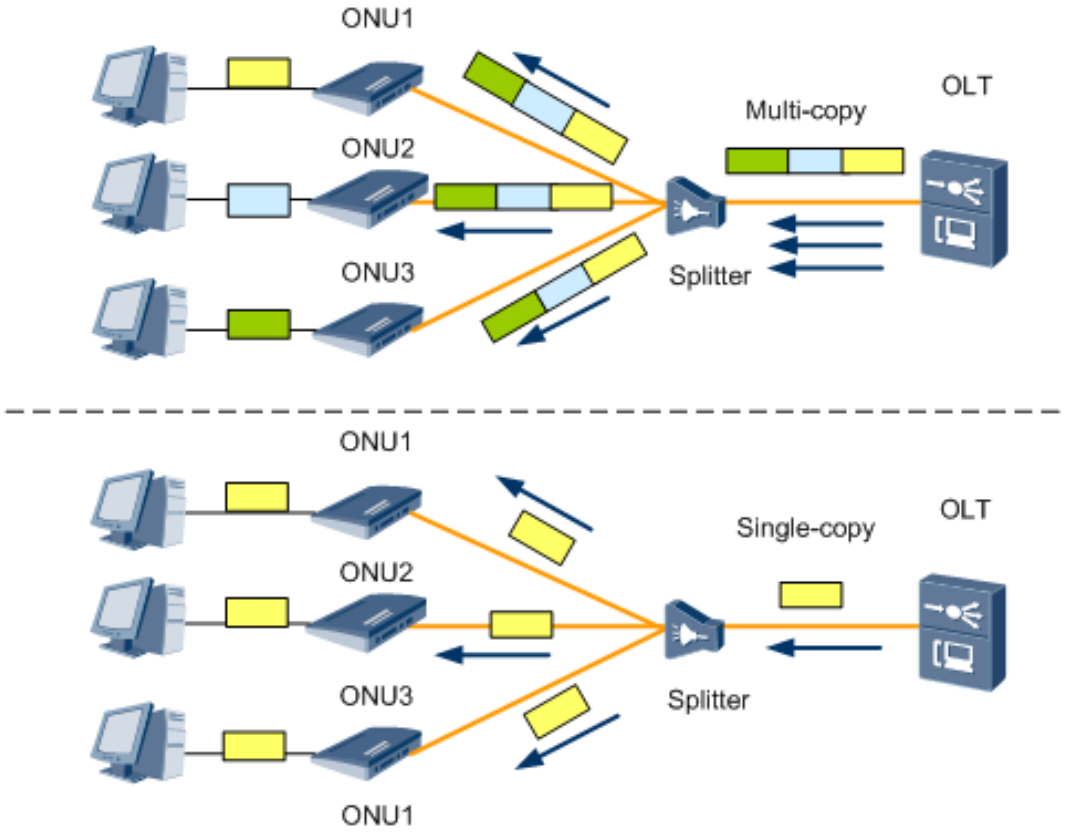
- Multi-copy duplication

  The primary difference between single-copy duplication and multi-copy duplication is that, in multi-copy duplication, multicast data is duplicated to corresponding service ports according to user requirements, encapsulated in the encrypted unicast GEM port channel, and then sent to the ONT. The following table shows the multicast forwarding table at the GPON board level.

Index	Entry
VLAN+GMAC	Multicast user list

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-33 Single-copy duplication and multi-copy duplication



The following table lists the differences between single-copy duplication and multi-copy duplication.

Item	Single-copy Duplication	Multi-copy Duplication
Duplication granularity	Based on GPON port	Based on multicast user
Bandwidth	One GPON port has only one multicast stream.	One multicast user has only one multicast stream, but one GPON port may have multiple multicast streams.
Security	On the one hand, the security depends on the ONT filtering; on the other hand, the head end and STB encryption system are required.	This mode uses the GPON line AES128 encryption system and the real-time key conversion function, which provides better security than the common encryption system of the head end.

 Copyright © Huawei Technologies Co., Ltd.

Item	Single-copy Duplication	Multi-copy Duplication
CAC of PON port	Supported	Not supported

13.4.4 Multicast Upstream Interoperation

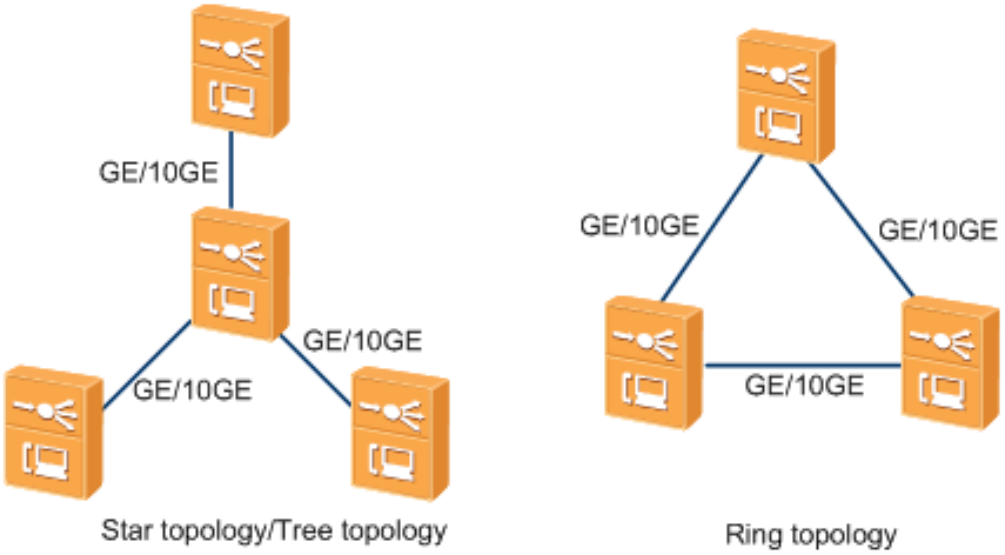
13.4.4.1 Multicast Cascading

- Ethernet cascading

  Using Ethernet cascading on the access device, the number of ports on the convergence device and the optical cable routing cost can be reduced. In addition, capacity expansion for more users in the residential community access area can be easily implemented.

  There are two common cascading network topologies, star (tree/chain) network and ring network, as shown in the following figure. Here, the star cascading network is used as an example. For details about the ring cascading network, see "**13.4.4.2 Ring Network of Uplink Ports**".

  Figure 13-34 Cascading network topologies

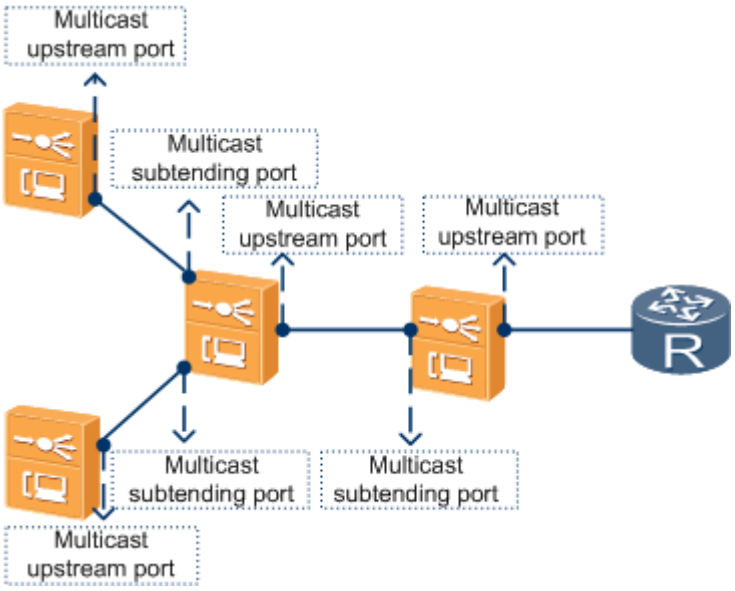


  – Multicast cascading port configuration

    The device, through the Ethernet port of its cascading boards, can be physically connected to the lower-layer device. Multicast service is configured through the multicast cascading ports, and in this way the interoperation between the devices is managed. A multicast cascading port corresponds to a physical port (the channel for carrying services can be created through the port VLAN or service stream). The following figure shows the relationship between the multicast cascading port and the multicast upstream port.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-35 Multicast cascading port and upstream port

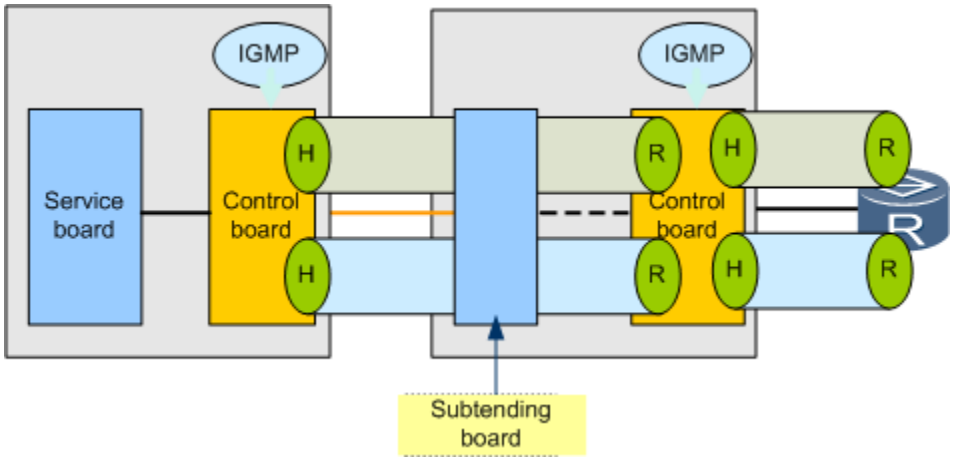


– IGMP control message

  In the multicast cascading scenario, the upper-layer device and the lower-layer device run the IGMP protocol stack separately. For a device, the cascading port (its lower-layer device) can be regarded as the multicast user. Multicast users are controlled by the lower-layer device and therefore the device does not support the following service functions for multicast users: rights management, multicast preview, multicast CAC, charging, and multicast service acceptance. The device supports the quick leave and normal leave functions.

  On the cascading port, the IGMP protocol stack is based on different VLANs, as shown in the following figure.

Figure 13-36 IGMP protocol stack on the cascading port



 Copyright © Huawei Technologies Co., Ltd.

NOTE

If an Ethernet port is not configured as the multicast cascading port, the Ethernet port discards the IGMP report message.

The (SIP, GIP) field of the IGMP message and the VLAN of the IGMP message are used for program matching. The policy of processing unmatched messages can be configured based on the cascading port.

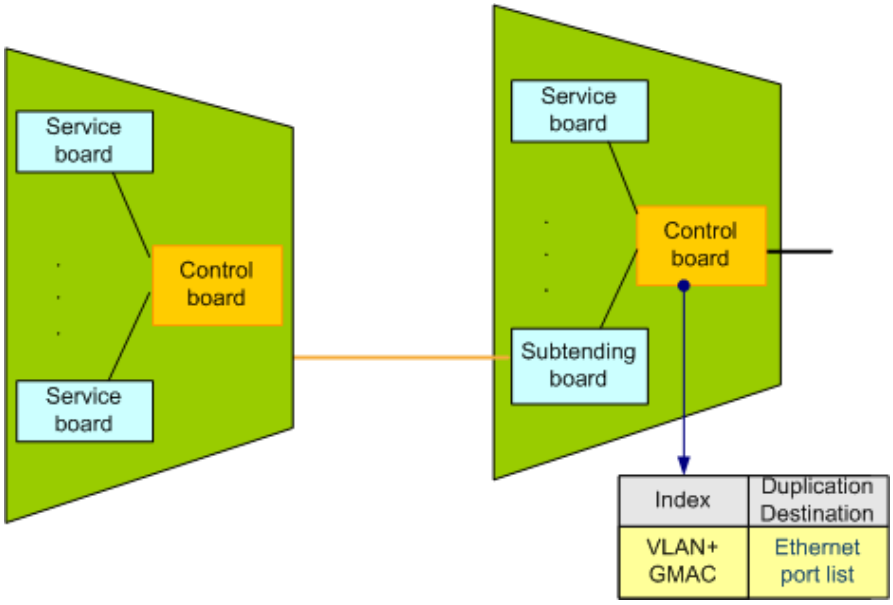
Considering the IGMP processing performance of the source node, it is recommended that all cascading devices adopt IGMP proxy instead of IGMP snooping.

– Multicast data forwarding

Multicast data can be forwarded only in a VLAN. According to different cascading boards, there are two forwarding architectures.

- One-level forwarding architecture: The following figure uses the control board as an example.

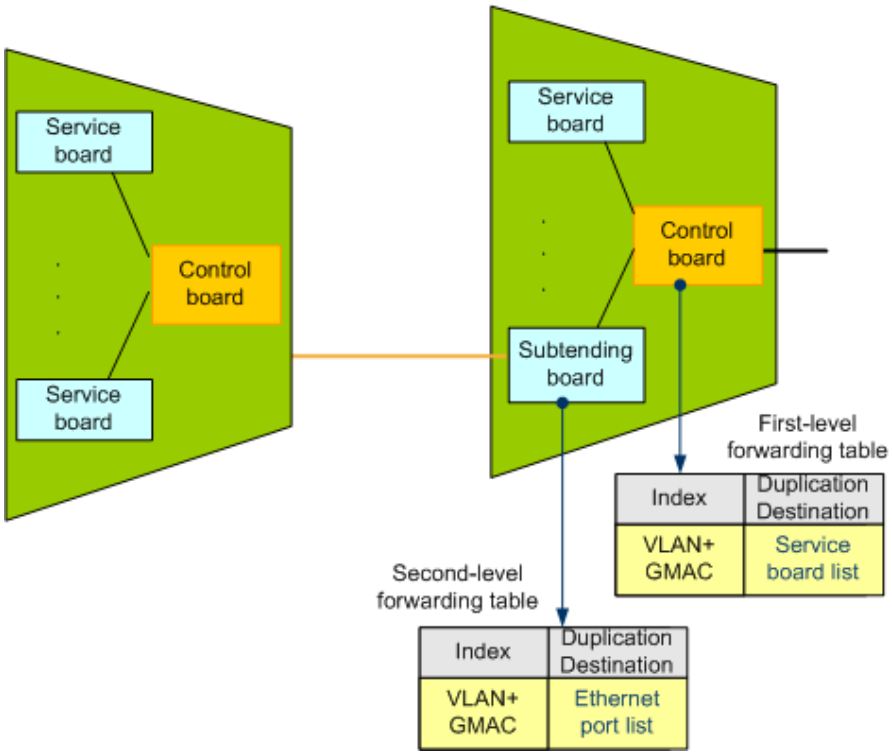
Figure 13-37 One-level forwarding architecture of multicast cascading



- Two-level forwarding architecture: The following figure uses the H901OXHD boards as an example.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-38 Two-level forwarding architecture of multicast cascading



- xPON cascading

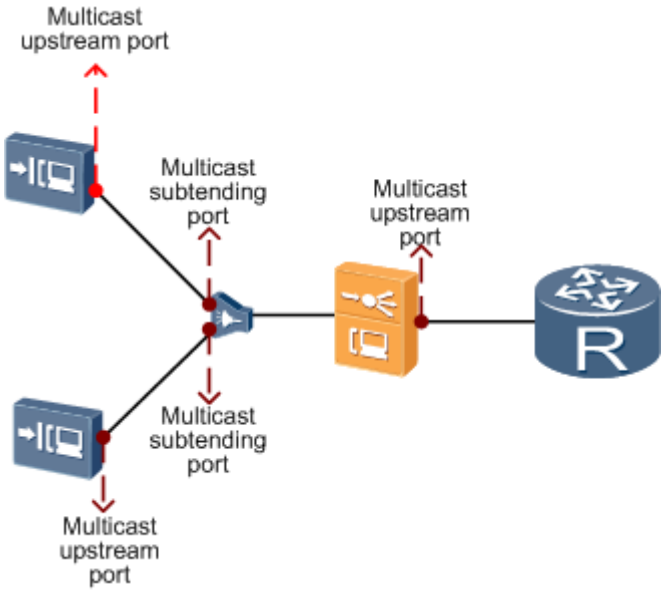
  xPON cascading meets the requirements for multicast services in the FTTC/FTTB scenario.

  – Multicast cascading port configuration

    The device can implement the physical connection between the OLT and the MxU by using the PON line. Similar to Ethernet cascading mode, in xPON cascading mode, the interconnection between devices is also managed through the multicast cascading port object. One xPON cascading port corresponds to a logical interface (GEM port or LLID). The actual bearer channel can be created by using the service port. The following figure shows the relationship between the xPON cascading port and upstream port.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-39 xPON cascading port and upstream port



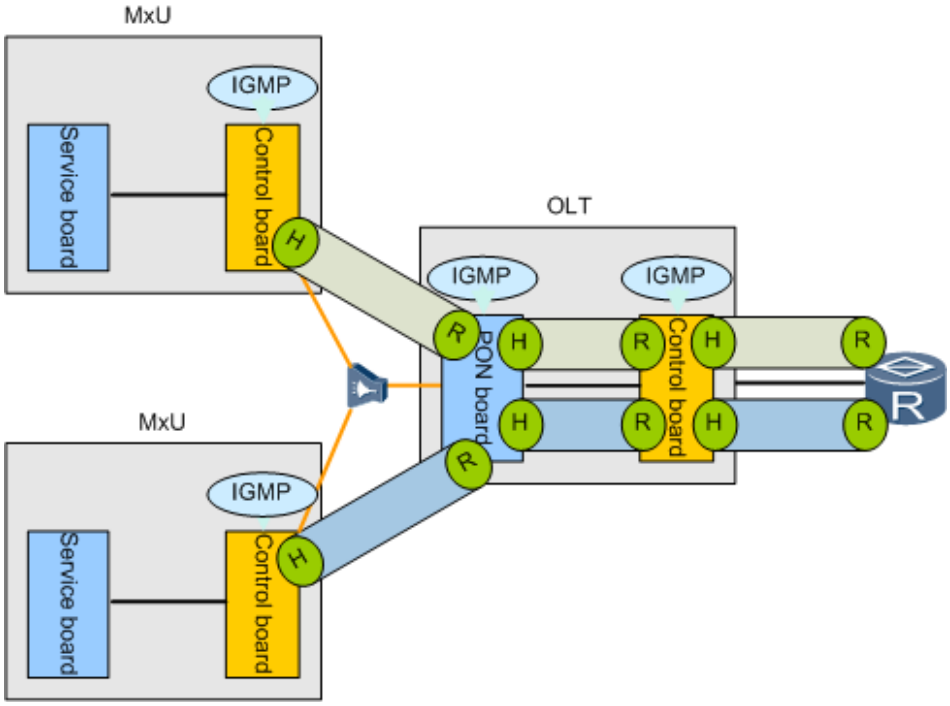
– IGMP control packet

  In the xPON cascading scenario, the OLT and the MxU run the IGMP protocol stack separately. Like Ethernet cascading, xPON cascading supports normal leave and quick leave.

  On the multicast cascading port, the IGMP protocol stack is based on different VLANs, as shown in the following figure. Different from Ethernet cascading, in xPON cascading, the bearer channel is limited by the maximum number of service ports that can be created on each GEM port or LLID, because the bearer channel is based on the service port. To support an MVLAN that is beyond the supported specifications, you can configure multiple GEM ports or LLIDs.

 Copyright © Huawei Technologies Co., Ltd.

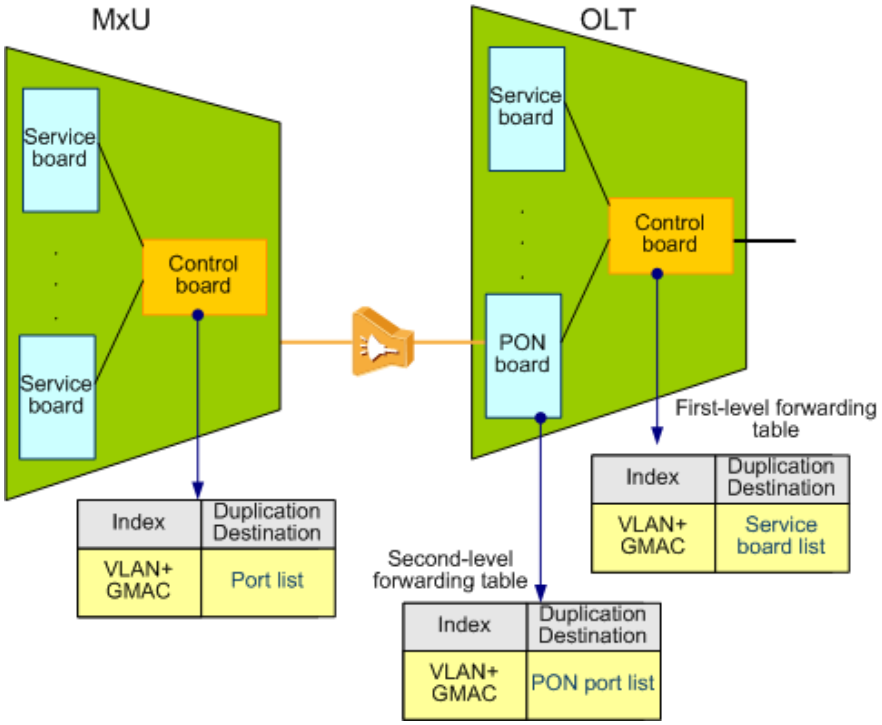
Figure 13-40 IGMP protocol stack of the xPON multicast cascading port



– Multicast data forwarding

  xPON cascading supports forwarding in the same VLAN and does not support cross-VLAN forwarding.

Figure 13-41 xPON multicast forwarding architecture



 Copyright © Huawei Technologies Co., Ltd.

NOTE

Service ports that adopt traffic classification by two-tagged VLANs do not support multicast cascading ports.

13.4.4.2 Ring Network of Uplink Ports

In the ring network, access devices on the physical link are connected to form a ring. Devices on the ring maintain the ring status by running the Layer 2 link protocol.

The ring network of access devices has two advantages:

- Low network construction costs: In the ring network, an access device does not need to connect to the convergence switch, but connects to its nearest access device. This significantly saves optical cable resources. The switch provides only a few ports for the access device. In the ring network, however, deploying a small number of switches can meet the access requirements.
- High reliability: The Layer 2 link protocol provides the uplink backup protection. With this function, when the uplink of a single access device is faulty, the device can switch to the backup uplink.

The multicast service supports the following ring networks on the network side.

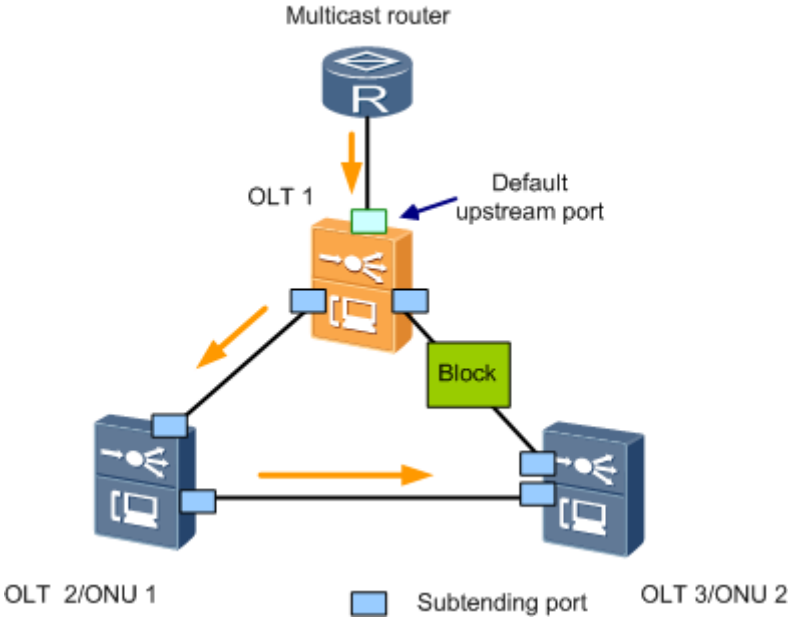
- MSTP

  The mode of the multicast uplink port needs to be set to MSTP by running the **igmp uplink-port-mode** command. In this case, the multicast uplink port of the device does not need to be configured; instead, the root port determined dynamically by MSTP serves as the multicast uplink port. If the access device is the MSTP root bridge (this device must be the injection point of the multicast data; using the MSTP priority configuration, ensure that this root bridge is not removed), the access device does not have the root port. Therefore, in the actual network, the multicast data injection port needs to be configured as the default multicast uplink port.

  In addition, the device ports on the ring need to be configured as the multicast cascading ports. The actual multicast downlink ports are determined by IGMP according to the multicast group membership table. The following figure shows the configuration of each role.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-42 Multicast configuration in the MSTP ring network

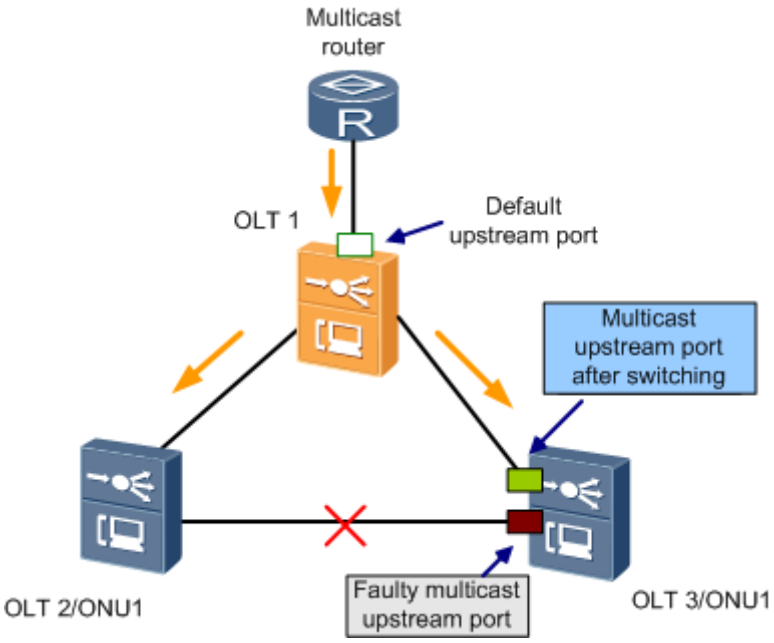


NOTE

The root bridge must be the injection point of the multicast service.

In the case of a link or device failure, after MSTP selects a backup link, the MVLAN-based IGMP protocol stack immediately sends the new root port (serving as the multicast uplink port) the join message targeting at the multicast group that the device is interested in. In this way, fast recovery of the multicast service can be ensured.

Figure 13-43 MSTP ring network fault



 Copyright © Huawei Technologies Co., Ltd.

- ERPS

  The mode of the multicast uplink port needs to be set to default by running the **igmp uplink-port-mode** command, and the multicast uplink port is configured as the video stream injection port. The master node of the ERPS ring must be planned in advance. The video stream injection node on the master node is the default uplink port. As shown in **Figure 13-44**, OLT1 is the master node of the ERPS ring, and P1 is the default uplink port.

  In addition, the node ports on the ring need to be configured as the uplink or cascading ports. The actual multicast downlink ports are determined by IGMP according to the multicast group membership table.

  NOTE

  The two ports on the master node on the ERPS ring must be configured as cascading ports, and ports connected to cascading ports on the master node must be configured as uplink ports. Other ports on the ERPS ring can be configured as uplink and cascading ports at the same time. As shown in **Figure 13-44**, P2 and P9 are cascading ports; P3 and P8 are uplink ports; other ports are mixed uplink and cascading ports.

  Then, the RPL owner port of the ERPS ring needs to be configured. As shown in **Figure 13-44**, P7 is the RPL owner port.

  NOTE

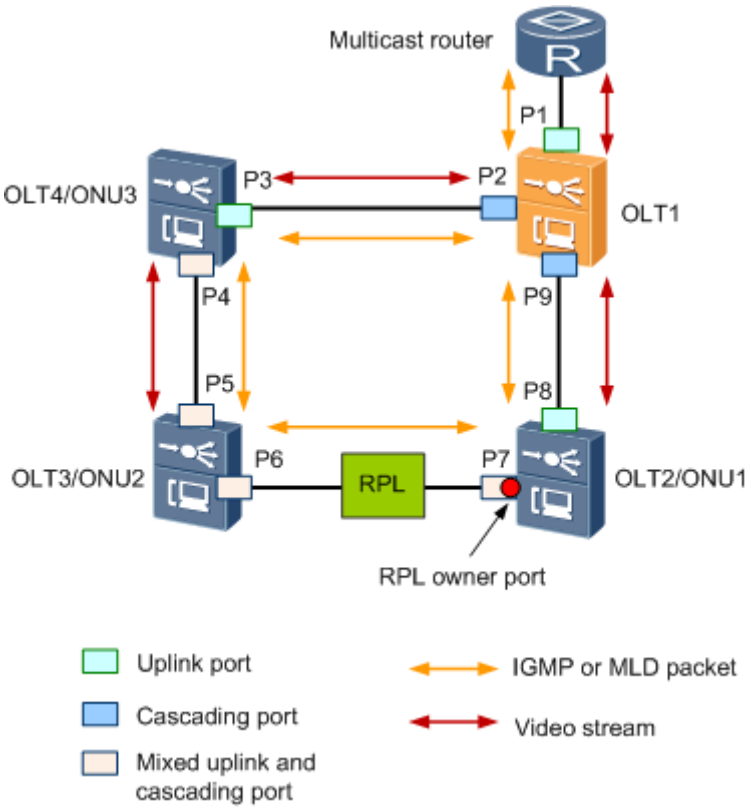
  One ERPS ring has only one RPL owner port, which is configured by the user. The RPL owner port can be blocked to prevent loops on the ERPS ring. A blocked RPL owner port cannot forward data packets but can send ring automatic protection switching (R-APS) packets.

  When the ERPS ring is carrying multicast services, you can run the **igmp ignore-block** command to set whether the RPL owner port of the ERPS ring can send IGMP or MLD packets. When the RPL owner port is enabled to send IGMP or MLD packets, dual back-up of multicast video data can be achieved. As shown in **Figure 13-44**, video stream data injected from the uplink port P1 is transmitted in two different directions to P2 and P9. Because P7 is blocked, the video stream transmitted from access node 1 to access node 2 is terminated at P7 and the video stream transmitted from access node 1 through access node 4 to access node 3 is terminated at P6. This ensures that each access node does not repeatedly receive data of the same multicast video stream.

  When link faults occur, the RPL owner port of the ERPS ring is switched to the forwarding status, and then the video stream on the RPL owner port is directly forwarded to the port that needs the video stream. If the RPL owner port is enabled to send IGMP or MLD packets, the status of the RPL owner port can be quickly switched during link faults.

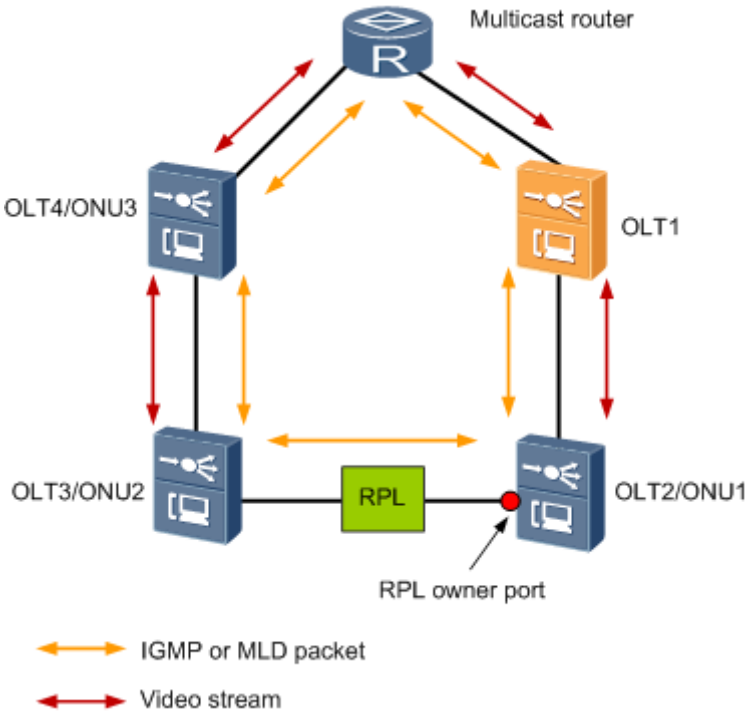
 Copyright © Huawei Technologies Co., Ltd.

Figure 13-44 Multicast configuration of the ERPS ring network



At present, the ERPS semi-ring network is frequently used, as shown in **Figure 13-45**.

Figure 13-45 ERPS semi-ring network application

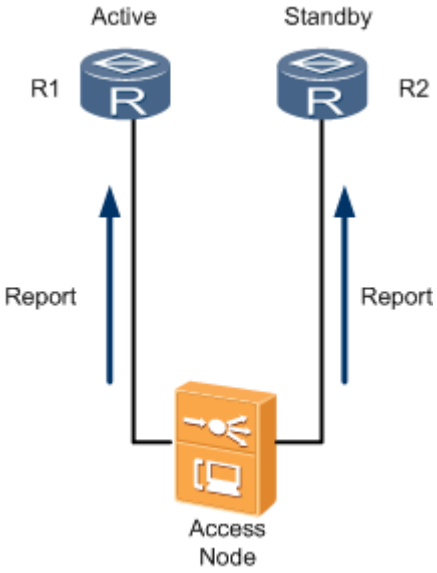


 Copyright © Huawei Technologies Co., Ltd.

13.4.4.3 Dual-homing of Upstream Ports

Multicast routers 1 and 2 function as the active router and standby router respectively, as shown in the following figure. To ensure fast recovery of the multicast service after a switching, use the IGMP message broadcast function provided by the access node.

Figure 13-46 Upstream port broadcasting IGMP messages



First, set the two access node ports connected to routers 1 and 2 as the multicast upstream ports (the two ports must not be in the same aggregation group or protect group). After this setting, when the access node transmits IGMP messages to router 1, it transmits the same IGMP messages to router 2 at the same time. In this way, router 2 can maintain in real time the same multicast forwarding entry as that of router 1. Once a switching occurs, router 2 can directly obtain the multicast forwarding entry and can ensure fast recovery of multicast service in a shorter time.

Note: If the router supports transfer of the multicast forwarding entry using a proprietary protocol, this can substitute for the upstream port dual-homing function. In this case, add the two access node ports to one aggregation group. Such a function is more commonly used in actual applications.

13.4.5 Advanced Multicast Technologies

13.4.5.1 Multicast Program Management

Multicast programs can be static programs or dynamic program.

Static Programs

A program list is configured before users can watch these video programs. In this way, controllable multicast is implemented using a rights profile. However, the

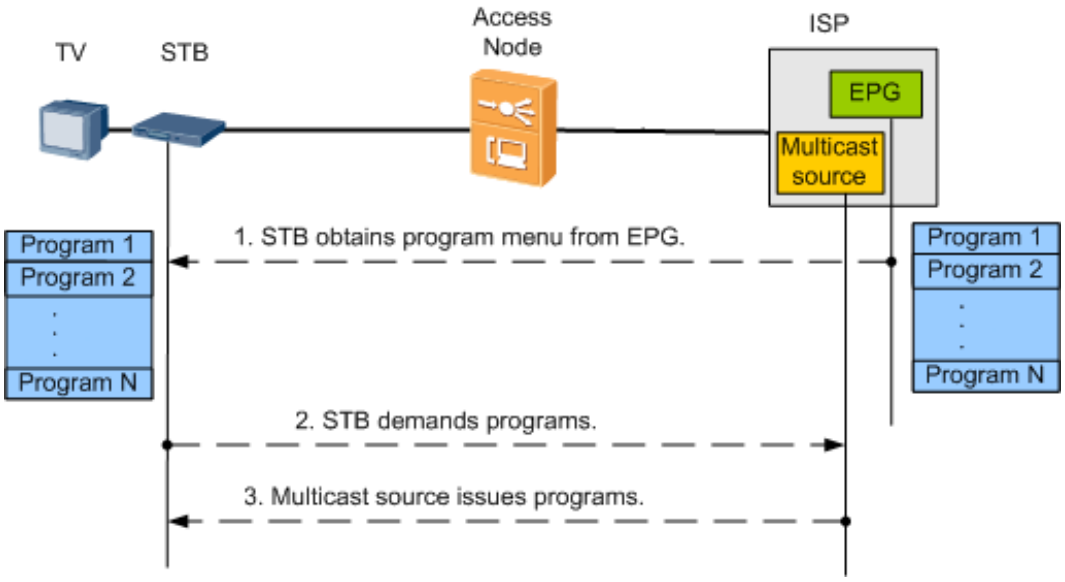
 Copyright © Huawei Technologies Co., Ltd.

program list and rights profile must be modified when the video service is modified. The device supports the program host, program prejoin, and multicast bandwidth management functions.

Dynamic Programs

In actual applications, if fine-grained management is not required on the device, dynamic programs can be applied. This avoids maintenance troubles brought by frequent program changes. In this case, program maintenance can be performed uniformly through the Electronic Program Guide (EPG) system.

Figure 13-47 Process of generating dynamic programs



1. After being started, the STB automatically obtains the program menu from the EPG server and provides the menu for the multicast user.
2. When the user orders a program, a corresponding IGMP message is generated and sent to the device. Hence, the program information on the device at this stage is not input by the administrator. Instead, it is dynamically generated in the MVLAN (to which the multicast user belongs) after the multicast group IP address and source IP address are extracted by the device from the real-time IGMP message of the multicast user.
3. The multicast program of the multicast source reaches the STB.

   To prevent the user from using an inappropriate group IP address, a legal multicast address segment can be configured based on MVLAN on the device for dynamic programs. According to the configuration, a multicast program is generated only when the group IP address is within the legal address segment; otherwise, the IGMP message of the user is dropped. Apart from the restriction by the address segment, the number of programs that can be dynamically generated is also controlled by hardware specifications.

   The fine-grained management that is not supported by dynamic programs on the device includes CAC, rights management, multicast preview, and pre-join.

 Copyright © Huawei Technologies Co., Ltd.

Hot programs are added as static programs and other programs are ordered by users dynamically. This configuration speeds up hot program ordering and shortens channel switching time.

13.4.5.2 Multicast Rights Management

Rights Management

With the method of configuring different multicast programs on different profiles, package-based rights management can be implemented on the device.

- Rights profile

  The rights to any multicast program can be specified in each rights profile, and each rights profile can be configured with a meaningful name. There are four types of rights:

  Forbidden: It indicates that a multicast user is not allowed to watch or preview a multicast program.

  Preview: It indicates that a multicast user can order a multicast program but is restricted in the watching duration and watching times.

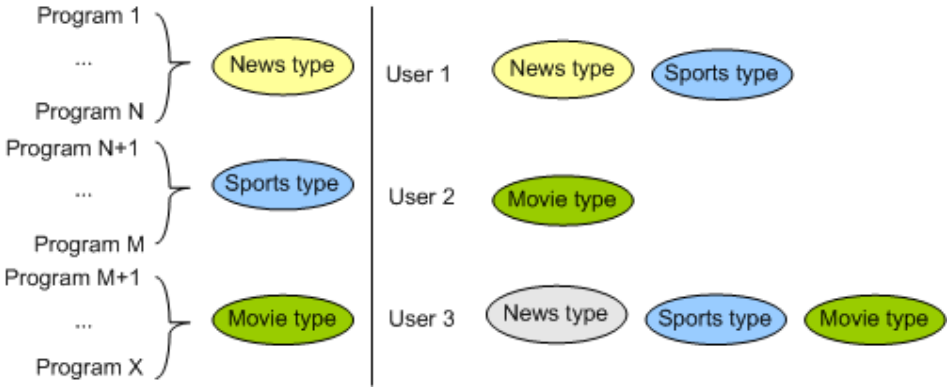
  Watch: It indicates that a multicast user can order a multicast program normally without any restriction.

  Idle: It indicates that a specific right is not assigned to a multicast program yet. It is the default value of the rights profile. The effect of "idle" equals that of "forbidden."

  Carriers can plan the rights profiles according to user-defined rules. Usually, there are three modes of planning.

  – The first one is planning by contents, such as planned as the news type, sports type, and movie type. In this case, one multicast program belongs to only one rights profile and the programs of different profiles do not overlap. Therefore, one user is usually bound to multiple profiles. See the following figure.

    Figure 13-48 Planning rights profiles-mode 1

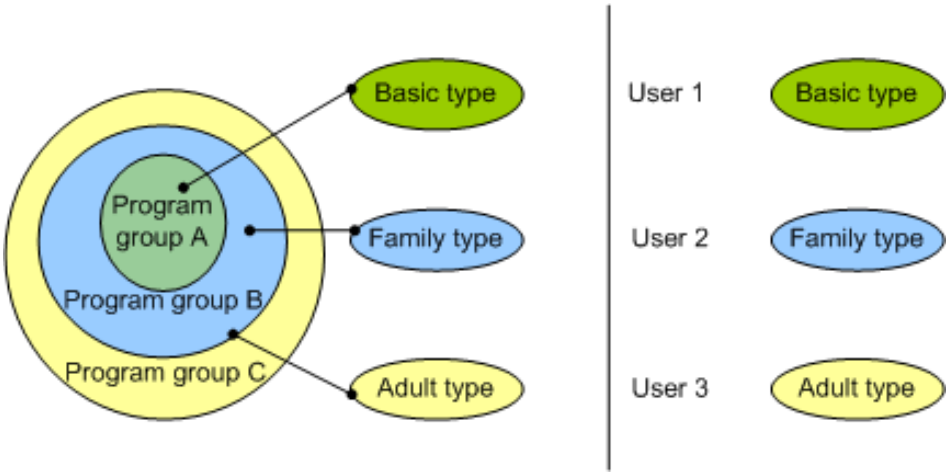
    

  – The second one is planning different levels by content volumes, such as planned as the basic type, family type, and adult type. In this case, one multicast program may belong to multiple rights profiles and the

 Copyright © Huawei Technologies Co., Ltd.

programs of different profiles may overlap. Therefore, one user is usually bound to only one profile. See the following figure.

Figure 13-49 Planning rights profiles-mode 2



– The third one is a hybrid of the first and second ones and is the most complicated as well as the most flexible mode. In this mode, the programs of different rights profiles may overlap and one user may be bound to multiple profiles. The same program may be configured with different rights in different rights profiles. To ensure that these rights profiles work as expected by the carrier when it comes to a specific program for a specific user, the rights of a program in the rights profiles must be prioritized. It is advised to plan the priorities before deployment to prevent any incorrect results. The following are examples.

Table 13-18 Priority of rights: forbidden > preview > watch > idle

Rights profile 1	Program 1: watch	User 1	Program 1: forbidden
Rights profile 2	Program 1: forbidden		

Table 13-19 Priority of rights: watch > preview > forbidden > idle

Rights profile 1	Program 1: watch	User 1	Program 1: watch
Rights profile 2	Program 1: forbidden		

- Rights control

  The rights of each multicast user can be configured by the following two steps:

 Copyright © Huawei Technologies Co., Ltd.

a. Plan the rights profiles of all multicast programs.
b. Bind a multicast user to the rights profiles required according to the contents subscribed to by the user.

   The device provides open MIB interfaces to support such operations.

   In addition, there is another method of implementing rights control: by configuring encryption on the head system and the STB. In this way, the carrier does not need to perform rights management on the device and only needs to enable or disable rights control at the system level or the multicast user level.

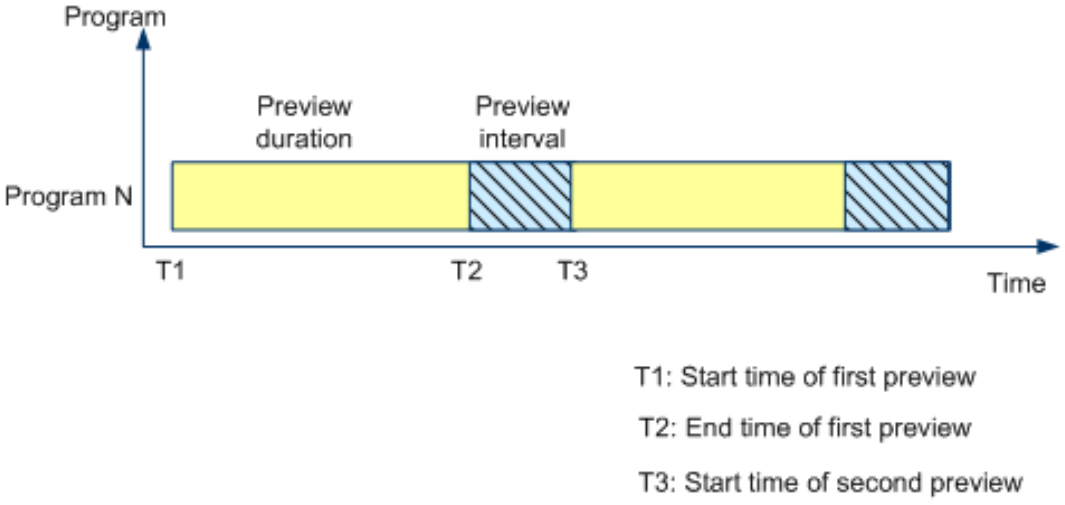
Multicast Preview

By providing the preview of certain special channels to multicast users, carriers may attract more users to subscribe to more programs with the watch right. Preview is usually used as an effective selling method.

The device manages the preview parameters of each multicast program by preview profiles. To be specific, each multicast program can be bound to a preview profile which is configured with preview parameters. Similar programs can be bound to the same preview profile to simplify management.

A preview profile contains three preview parameters.

Figure 13-50 Preview parameters



- Preview interval: It is the minimum interval between two previews. The interval is from the end time of the previous preview to the start time of the current preview (from T2 to T3 as shown in the preceding figure). If the interval between the two previews of a user does not reach the specified preview interval, the user is currently not allowed to preview the program. Such a mechanism guards against any "rogue" behavior of users. Without the restriction of the preview interval, a user may keep previewing the same program and is actually "watching" a program without having to pay for the "watch" right.
- Preview times: This parameter specifies how many times a multicast user is allowed to preview the same program during a day. Each time the user leaves

 Copyright © Huawei Technologies Co., Ltd.

a previewed program, the counter increases by 1. When the counter exceeds the maximum value, the further orders of the user for the program will be rejected. In this case, the user's right to the program can be regarded as demoted to "forbidden." However, the preview right can recover the next day.

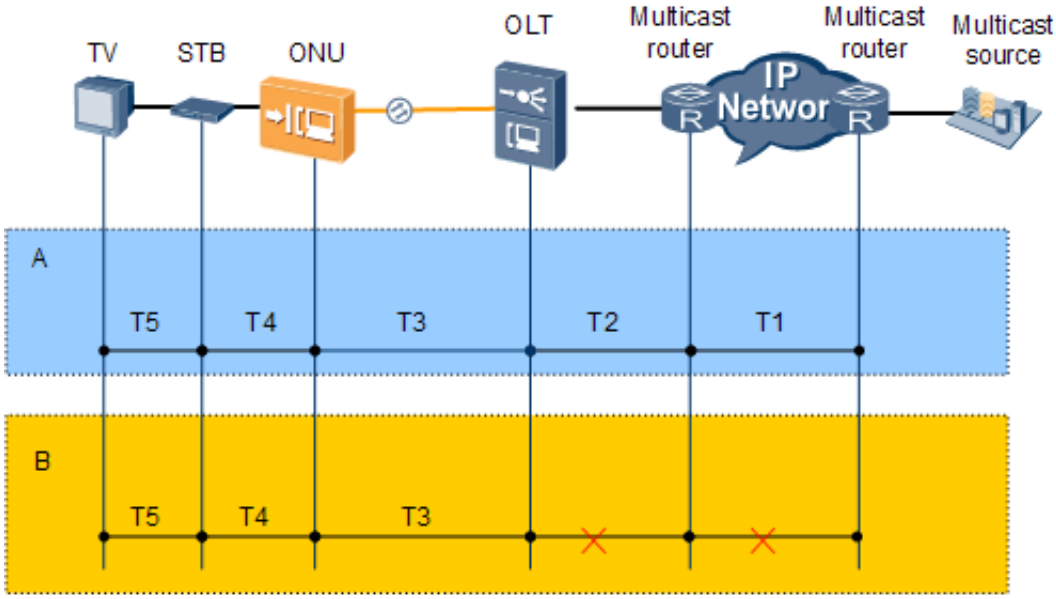
- Preview duration: This parameter specifies for how long a multicast user is allowed to watch the same program each time. The duration starts from the beginning of the order (from T1 to T2 as shown in the preceding figure). After the duration expires, the user will not be able to receive any data for the multicast program.

For details on how to control the preview of multicast users, see "**Rights Management**."

Prejoining a Program

The prejoin function is used to shorten the course of channel switching (reduce switching latency), so as to improve users' experience in channel switching. Switching latency comprises the processing consumption in each segment of a network, as shown in the following figure. With the prejoin function enabled, the network-side processing consumption (T1+T2) equals 0.

Figure 13-51 E2E multicast switching latency



A: Processing latency of non-prejoin program = T1 + T2 + T3 + T4 + T5

B: Processing latency of prejoin program = T3 + T4 + T5

The prejoin function applies to the IGMP proxy scenario. It equals that there is always online users for a program.

- The flow of prejoining a program is the same as the flow of joining a program normally. For details, see section 1.6.4. Once a multicast stream is successfully demanded, it is transmitted to the DSLAM.

 Copyright © Huawei Technologies Co., Ltd.

- In the flow of leaving a prejoined program, compared with the flow of leaving a normally joined program, the DSLAM does not transmit the leave message to the multicast router even when the last multicast user leaves the program.
- In the flow of querying a prejoined program, compared with the flow of querying a normally joined program, the DSLAM responds to the multicast router's query as required by the protocol regardless of whether or not the multicast group membership table of the program contains a multicast user.

From above all, viewed from the router, there are always online users for a prejoined program.

The prejoin function can be set for a program. In general, set the prejoin function for the program that is most commonly demanded by users. A dynamic program does not support the prejoin function.

13.4.5.3 Multicast CAC

CAC is the short form for call admission control. Here, it means controlling the setup of IGMP sessions. If an IGMP session fails to be set up, a multicast user will fail to receive the multicast program ordered.

In a broad sense, implementing CAC requires implementing the first-level control in the system. Currently, system control includes the following:

- Anti-DoS attack. The rate of IGMP messages sent from the user side must not exceed the specified value in the system. Otherwise, the system will regard that a DoS attack occurs and drops the messages. Such a protection method applies not only to IGMP messages, but also to control packets such as DHCP and PPPoE packets. For details, see "Anti-DoS Attack."
- Anti-IP spoofing. When this function is enabled, the user must obtain a legal IP address through DHCP before ordering any program. Only the IGMP messages using the legal IP address as their source IP address will be accepted by the system; otherwise, the messages will be regarded as coming from unauthorized users and will be dropped by the system. For details, see "Anti-IP Spoofing."

  NOTE

  Only the centrally-controlled multicast supports this feature.
- Broadband message overload. When a service traffic burst occurs, the system resources may not be able to support all services. Then, the system will drop certain messages according to specified policies to ensure that the services with a higher priority are not affected. In this case, IGMP messages may be "sacrificed" to reduce the system load. For details, see "Broadband Message Overload."

  After the first-level control in the system comes the multicast first-level control, which includes the following:
- Concurrent number of programs of a multicast user. This parameter specifies how many channels a multicast user is allowed to order at the same time. The parameter can be configured based on multicast user.
- Rights control. For details, see "**13.4.5.2 Multicast Rights Management**."
- Bandwidth check. Though the system supports QoS control on various types of traffic, packet loss (drop by priority or tail drop) may still occur when the

 Copyright © Huawei Technologies Co., Ltd.

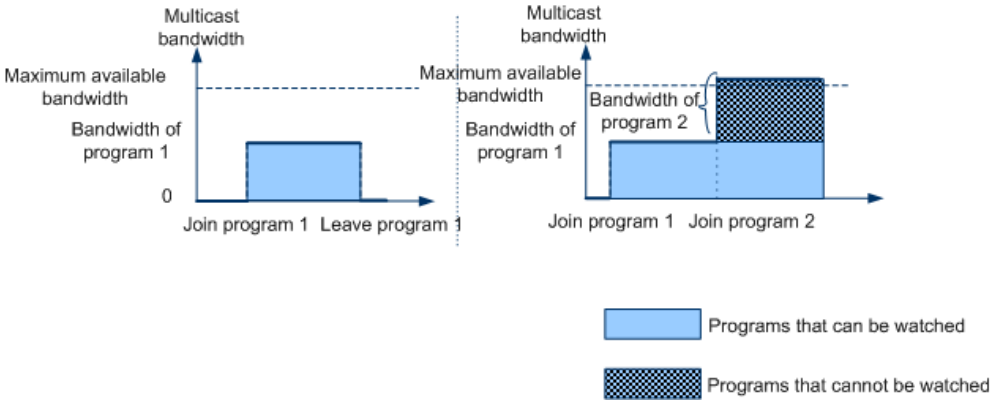
transmission bandwidth is overloaded. However, due to the real-time and non-retransmittable properties of multicast programs, postmortem QoS will directly cause pixelation to the programs with packet loss (not only to newly ordered programs). Hence, the requirements of IPTV for high-quality experience are not met. Bandwidth check enables the system to control a newly ordered program beforehand. In this way, the system can ensure that the programs that have been ordered enjoy sufficient bandwidth and will not be affected by the new program. With bandwidth check, only the newly ordered program is affected (if bandwidth is insufficient, the user will not be able to watch the newly ordered program).

CAC can be classified into the following types according to different control points and methods.

- Multicast user bandwidth CAC

  First, each pre-configured program is configured with bandwidth. The bandwidth is configured with reference to the video bit streams, and the margin of packet encapsulation and network transmission jitter; if possible, actually tested network traffic can also be used, as a better reference. Then, each multicast user is configured with available bandwidth. The available bandwidth is configured with reference to the actual line bandwidth or the planning of service provisioning. Hence, when receiving the first IGMP join message of a program, the device subtracts the bandwidth occupied by the program from the available bandwidth of the user. If the remainder is smaller than 0, the device rejects the order request of the user. When receiving an IGMP leave message of a program, the device returns the bandwidth occupied by the program to the available bandwidth of the user. The time of returning is the time when the device stops forwarding multicast data. That is, the program is not ordered by any end user of the terminal.

  Figure 13-52 Multicast user bandwidth CAC

  Multicast user bandwidth CAC can be configured at the system level or at the multicast user level.

- GPON port bandwidth CAC

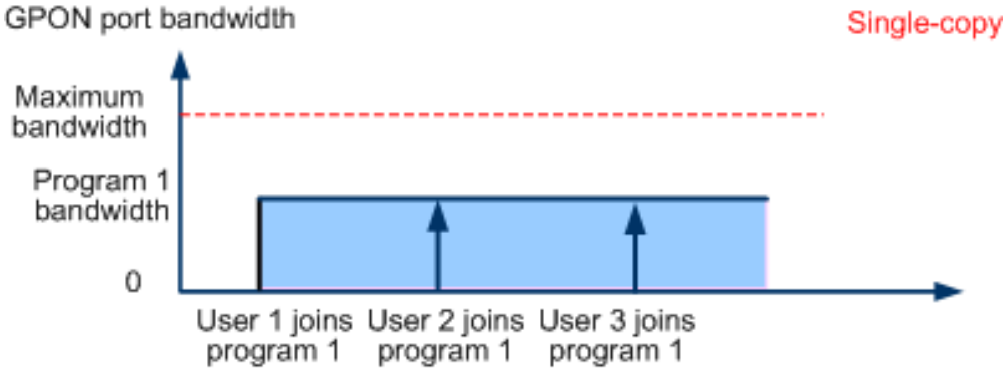
  GPON single-copy duplication function (default configuration): Under the same GPON port, even if multiple multicast users order the same multicast program, the multicast data is duplicated only once and sent to corresponding multicast users through the downstream multicast channel. Therefore, this

 Copyright © Huawei Technologies Co., Ltd.

function ensures that the downstream multicast bandwidth does not overflow the downstream line bandwidth of the GPON port.

To do so, the operator first needs to configure bandwidth for each pre-configured program (see "Multicast user bandwidth CAC"), and then allocates the available bandwidth for each GPON port (depending on the actual line bandwidth or the service provisioning plan). In this way, after receiving the first IGMP join message of the program, the device deducts the bandwidth of the corresponding program from the remaining bandwidth of the GPON port. If the deduction result is smaller than 0, the device rejects the order of the user. After receiving an IGMP leave message, the device returns the bandwidth of the corresponding program to the GPON port (the moment of returning is when the forwarding of multicast data is stopped, that is, no multicast user under the GPON port orders this program).

Figure 13-53 GPON port bandwidth CAC



This function can be configured at the system level or GPON port level. It can be used together with the multicast user bandwidth CAC.

13.4.5.4 Multicast QoS

This topic describes the quality of service (QoS) features dedicated to the multicast service. For more information about QoS (including traffic classification, traffic policing, ACL policy, and congestion avoidance and management), see the **QoS** feature.

Priority Processing of IGMP Packets

The device supports processing of only the 802.1p priority of IGMP packets.

 Copyright © Huawei Technologies Co., Ltd.

Table 13-20 Priority processing in IGMP proxy/snooping

Cascading Mode	Upstream	Downstream
Multicast user/xPON cascading	Based on MVLAN.	Traffic classification methods: • VLAN: by the 802.1p priority specified in the traffic profile. • VLAN+encapsulation type: by the 802.1p priority specified in the traffic profile. • VLAN+802.1p priority: by the 802.1p priority specified by traffic classification.
Ethernet cascading	Based on MVLAN.	Based on MVLAN.

Priority Processing of Multicast Traffic Streams

The device supports processing of only the 802.1p priority (CoS) of the multicast traffic streams.

Table 13-21 Downstream multicast priority processing

Cascading Mode	Pre-configured Program	Dynamic Program
Multicast user/xPON cascading	Traffic classification methods: • VLAN: by the 802.1p priority specified in the traffic profile. • VLAN+encapsulation type: by the 802.1p priority specified in the traffic profile. • VLAN+802.1p priority: by the 802.1p priority specified by traffic classification.	
Ethernet cascading	Based on program.	The priority is not configurable. That is, the priority carried remains unchanged.

13.4.5.5 GPON ONT Multicast

The GPON end-to-end multicast service requires the cooperation of the ONT. The following points must be noted:

- VLAN translation

  If the carrier plans the home gateway at the user's house, generally, the VLAN of the IPTV service (also called C-VLAN) needs to be planned. Because the OLT does not directly support translation of the MVLAN, the operator can configure VLAN translation on the ONT to meet the planning requirement

 Copyright © Huawei Technologies Co., Ltd.

(the OLT provides the corresponding CLI and the configuration can be issued to the ONT through OMCI). The MVLAN can be translated in three ways: transparently transmitted, translated to untagged, and translated to a specified VLAN.

- Controllable multicast

  In single-copy duplication, GPON downstream multicast programs are broadcast. Assume the following condition: After an authorized multicast user orders a program, all users under the GPON port can receive this program. Therefore, to implement complete rights control on the access device, the OLT must configure the ONT to work in the "dynamic controllable" mode. In this way, the multicast filtering table (white list) on the ONT is issued by the OLT after multicast control checking. If a downstream multicast program is not in the multicast filtering table, the ONT cannot receive this multicast program.

  If the ONT is configured to work in the "IGMP snooping" mode, the multicast filtering table on the ONT is completely maintained by the ONT. In this case, multicast program rights management is generally implemented by the encryption system of the IPTV platform.

13.4.5.6 Multi-instance Multicast

With the increased use of open networks, carriers' networks need to provide independent multicast domains for different multicast ISPs so that different ISPs do not interfere with each other. Independent multicast domains can be implemented on the management plane, control plane, and forwarding plane by planning different MVLANs on the device.

- Management plane (ASMSSM or ASM ONLY)

  Within each MVLAN, the multicast programs to be provisioned, and the multicast upstream ports and multicast users involved can be configured for each ISP. Here, the multicast programs need to be noted. To ensure that each ISP can plan multicast programs independently, the multicast program triplet (MVLAN, source IP address, and multicast IP address) needs to observe the following rules:

  – If two GIPs are mapped to the same GMAC (for details on the mapping method, see "**Basic Concepts**"), the two GIPs are regarded as the same GIP.

    NOTE

    If the device does not have boards that support only VLAN+GMAC forwarding chip, this rule may not be considered; otherwise, carriers need to follow this rule in planning.

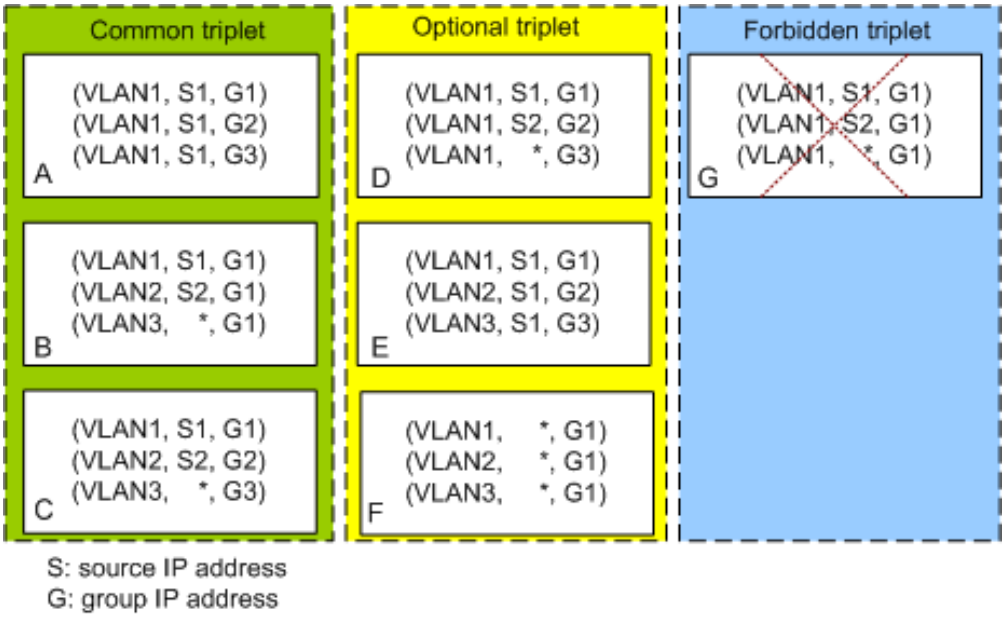
  – To ensure the uniqueness of a multicast forwarding entry on the forwarding plane, (MVLAN, multicast IP address) must be unique.

  – Especially, in the case of an IGMPv2 message or an IGMPv3 message in the ASM mode, the multicast source IP address equals any value (usually represented as * or as any). In this case, only the second rule needs to be observed.

  Use section G in the following figure as an example. According to the second rule, (MVLAN, multicast IP address) must be unique, but (VLAN1, G1) in section G is not unique. Therefore, configuring or generating the entries in section G is not allowed. The entries in the other sections in the following figure can also be judged by the rules described above.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-54 Multicast program triplet (ASMSSM or ASM ONLY)



- Management plane (SSM ONLY)

  Within each MVLAN, the multicast programs to be provisioned, and the multicast upstream ports and multicast users involved can be configured for each ISP. Here, the multicast programs need to be noted. To ensure that each ISP can plan multicast programs independently, the multicast program triplet (MVLAN, source IP address, and multicast IP address) needs to meet the following rules:

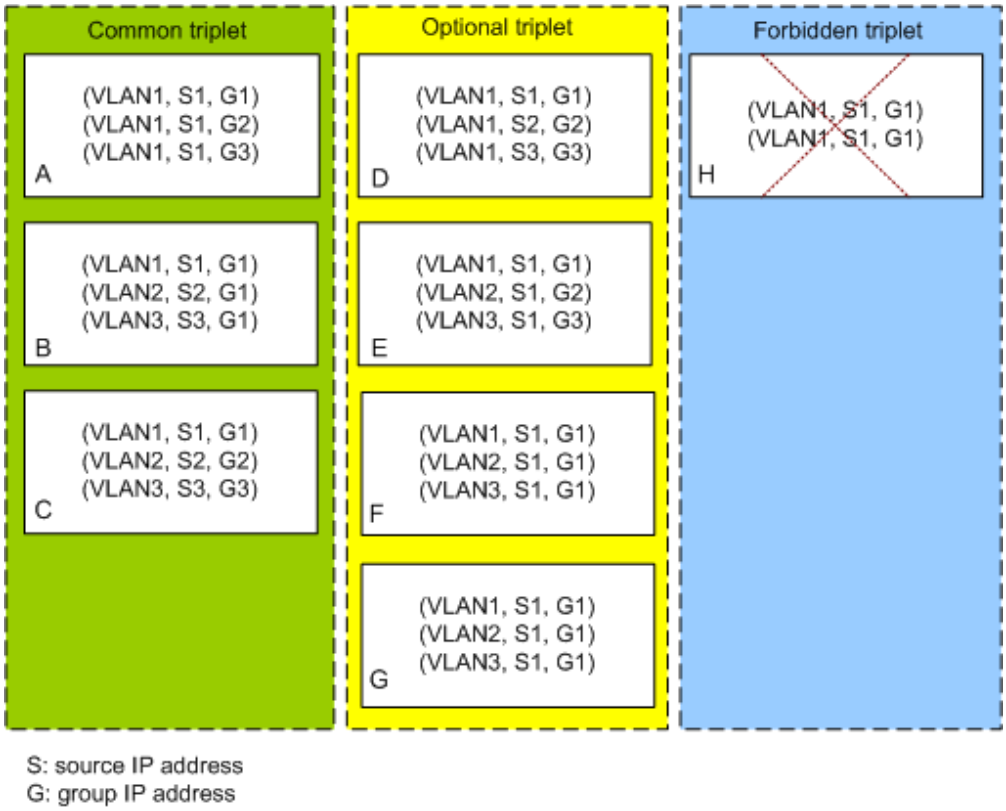
  – If two source IP addresses have the same least significant 20 bits, these two source IP addresses are considered the same source IP address. Carriers need to have proper planning to avoid this.

  – To ensure the uniqueness of a multicast forwarding entry on the forwarding plane, (MVLAN, multicast IP address, and source IP address) must be unique.

  Use section H in the following figure as an example. According to the second rule, (MVLAN, multicast IP address, and source IP address) must be unique, but (VLAN1, G1, and S1) in section G is not unique. Therefore, the entries in section H are not allowed to be configured or generated. The entries in the other sections in the following figure can also be judged by the rules described above.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-55 Multicast program triplet (SSM ONLY)



- Control plane

  On the network side, each MVLAN has an independent IGMP protocol stack. Each ISP can select the protocol version, message priority, and IGMP proxy or IGMP snooping. On the user side, each multicast user has an independent IGMP protocol stack and is not affected by other multicast users. **Table 13-22** describes the IGMP message processing policies in different group filter modes.

  Table 13-22 IGMP message processing policies

Group Filter Mode	IGMPv2	IGMPv3 Without Source	IGMPv3 with Source
ASMSSM	Processed	Processed	Processed
ASM ONLY	Processed	Processed	Dropped
SSM ONLY	Dropped	Dropped	Processed

- Forwarding plane

  On the forwarding plane, multicast forwarding tables with the group filter mode use ASMSSM or ASM ONLY use the MVLAN and multicast IP address together as indexes (VLAN+GMAC as the index for boards that do not support VLAN+GIP). This ensures that different MVLANs do not interfere with each other. Multicast forwarding tables with the group filter mode SSM ONLY use

 Copyright © Huawei Technologies Co., Ltd.

MVLAN, multicast IP address, and source IP address together as indexes. For details, see "**Multicast Forwarding Table**." For the control board, implementing QoS scheduling on the traffic of different MVLANs on the same port on the service board equals unicasting. For details, see "QoS".

NOTICE

1. In most situations, the MVLAN is different from the S-VLAN of the traffic stream to which the multicast user belongs. This S-VLAN can be used in the IPTV VoD service.
2. VLANs in S-VLAN+C-VLAN forwarding mode are used for transmission over point-to-point channels and are not applicable to multicast duplication and forwarding. Therefore, the S-VLAN in S-VLAN+C-VLAN forwarding mode cannot be used as an MVLAN.
3. When the group filter mode is ASMSSM, binding a multicast user to multiple MVLANs is not recommended. If a home has subscribed to multiple ISP programs, configure different multicast users for different MVLANs to ensure that one-to-one mapping between MVLANs and multicast users.
4. When the group filter mode is ASM ONLY or SSM ONLY, binding a multicast user to multiple MVLANs is not supported.

13.4.5.7 Charging Mode

For multicast services, carries or ISPs usually adopt two charging modes:

- Fixed charging. In this mode, programs are divided into different packages. The user needs to pay a fixed amount of fee for each package in a fixed period (such as by the year or by the month). This charging mode does not restrict the multicast user in the order count or the ordered volume of traffic.
- Pay per view (PPV): In this mode, the user is charged according to the order count of different programs.

In the first charging mode, since it is fixed charging, the charging does not relate to the behavior of the multicast user. Therefore, the first mode is supported by the device inherently and does not require additional functions from the device.

In the second mode, the device records the order behavior of each multicast user or each multicast terminal identified by the MAC address, and provides the behavior information in the form of a call detail record (CDR) to the accounting system for settling an account. The complete configuration of the CDR function consists of three steps:

1. Enabling the logging function. The function can be configured at the multicast user level, multicast program level (configurable for pre-configured programs, and defaulted to **enable** for dynamic programs), or system level. When a user finishes a complete watch behavior (from the program order starting to ending), or when the user fails to order a program because of failing to pass the multicast CAC, a log is generated.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

When the logs reach the maximum capacity, new logs will overwrite old ones. Therefore, to prevent heavy consumption of log resources in the case where the user quickly browses through channels, the device supports a configurable flag time for generating logs. If the multicast user watches a channel for a duration shorter than the flag time, the device does not generate a log. On the contrary, to timely log the users who stay online for a long time, the device supports the configuration of another time value (if the value is 0, the log will not be generated). When a user stays online for longer than the preset value, the device automatically generates a log.

2. Configuring the file server. The operator needs to select a CDR transfer file. Available options are TFTP/FTP/SFTP. Also, the operator needs to set the IP addresses of the primary and secondary servers.

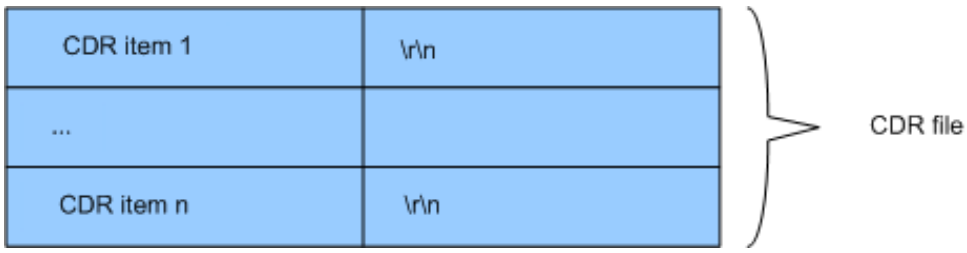
NOTE

FTP and TFTP have security risks due to protocol restrictions. SFTP is recommended.

3. Enabling the CDR functions (at system level). After the CDR function is enabled, the device automatically integrates the logs that need to be reported into a text file and transfers the file to the server when either of the following conditions is met: when the reporting interval expires, or when the number of logs reaches the reporting threshold.

   There are two types of CDR but they have the same format of the text file, as showed in **Figure 13-56**. The format of the text file name is different from the detailed format of a CRD item.

Figure 13-56 Format of the text file



– Recording the order behavior of each multicast user

  The format of the text file name is **HWCDR**-*host name*-*YYYYMMDDHHMMSS*.**txt**.

Table 13-23 Detailed format of a CDR item

ID	Field Name	Specification	Commentary
0	TAG	3 Bytes	Fixed as "Log". "Log" is the module name which generates syslog
1	SN	0..5 Bytes	Using a 16 bit variable to record. The maximum value is "65535" which occupies 5 bytes.

 Copyright © Huawei Technologies Co., Ltd.

ID	Field Name	Specification	Commentary
2	FrameSlotPortGemport	5..13 Bytes	F/S/P/GemPort for GPON user
	FrameSlotPortFlow	5..14 Bytes	F/S/P/FlowID for other type user
3	ProgramIP	0..15 Bytes	Sample: 239.1.1.1
4	OperMode	0..1 Bytes	0-Watch; 1-Preview; 2-No Right; Other is invalid
5	StartDate	0..18 Bytes	YYYY-MM-DD HH:MM:SS
6	EndDate	0..18 Bytes	YYYY-MM-DD HH:MM:SS
7	ProgramName	0..16 Bytes	Sample: cctv1. If the program does not exist, this parameter is "No-Name".
8	ProgramSrcIP	0..15 Bytes	Sample: 192.168.1.1. If the IP address is invalid, this parameter is "*".
9	Reason	1..2 Bytes	Syslog generation reason: 11: User's online time is too long 0: User leave.

– Recording the order behavior of each multicast user or each multicast terminal identified by the MAC address

  The format of the text file name is **BTVCDR**-*host name*-*X*-*YYYYMMDDHHMMSS****:txt**.

  NOTE

  When the value of X is A, automatic reporting is enabled. When the value of X is M, manual reporting is required. *** indicates millisecond.

 Copyright © Huawei Technologies Co., Ltd.

Table 13-24 Detailed format of a CDR item

ID	Field Name	Specification	Commentary
1	Record type	1..2 Bytes	• 0: Successful channel zap • 1: Channel timeout by general query or group-specific query • 2: Successful preview channel zap • 3: Preview channel timeout by general query or group-specific query • 4: It is automatically generated every N hours even when there is no channel (including preview channel) zap. • 5: Join a channel with no access right • 6: Join a channel with preview access right (but with preview limit exceeded)
2	System name	1..50 Bytes	The name of the OLT/DSLAM.
3	IGMP user	7..14 Bytes	Frame ID/Slot ID/Port ID/Service port Index Sample: 0/1/1/233
4	Terminal IP	7..15 Bytes	Zero-length string means IP is unknown. Sample: 192.168.0.1
5	Terminal MAC	17 Bytes	Zero-length string means MAC is unknown. Sample: 00-e0-fc-b4-0b-ee
6	Multicast VLAN	1..4 Bytes	Multicast VLAN of the program Sample: 200
7	Program Group IP	9..15 Bytes	Zero-length string means source IP of the program is not configured. Sample: 10.1.1.1
8	Program Source IP	0..15 Bytes	Sample: 192.168.1.1. If the IP address is invalid, this parameter is "*".

 Copyright © Huawei Technologies Co., Ltd.

ID	Field Name	Specification	Commentary
9	Start viewing time	19 Bytes	YYYY-MM-DD HH:MM:SS{+|-}hh:mm [DST]. Sample: 2011-08-27 09:30:20+08:00
10	Time of record generation	19 Bytes	YYYY-MM-DD HH:MM:SS{+|-}hh:mm [DST] Sample: 2011-08-27 10:30:20+08:00
11	Program Name	1..16 Bytes	Sample: cctv1
12	Duration viewed	1..n Bytes	Viewing duration in seconds. Sample: 30

13.4.5.8 Double-VLAN Tag Multicast

Double-tag multicast specially refers to the number of VLAN tags of packets carried by the network-side multicast of the access device. In actual multicast networks, generally only one VLAN tag is used for multicast. The reasons for using double-tag multicast are as follows:

- The unicast application is limited by VLAN number supported by device and more and more applications adopt the double-tag planning. In addition, because the convergence switches of some vendors interconnected with the access device do not support transmission of single-tag packets and double-tag packets on the same physical link, the multicast has to adopt the double-tag mode.
- To use the VLAN planning like that of unicast in a unified manner. For example, outer VLAN tags indicate different ISPs and inner VLAN tags indicate different services.

The device can be configured whether to use the double-tag multicast function.

- IGMP control message

  Double-tag multicast and single-tag multicast have the same processing flow for IGMP control messages (see "**Join Flow**"). Difference between them: On the network side, the transmitted and received IGMP messages in single-tag multicast have one tag whereas the transmitted and received IGMP messages in double-tag multicast have two tags The outer VLAN tag in double-tag multicast is the MVLAN to which the program belongs and the inner VLAN tag can be configured based on MVLAN (adopting the "easy in strict out" principle, the device does not check the inner tag and configuration consistency of received IGMP messages). The inner tag priority and outer tag priority of IGMP packets can be configured separately based on MVLAN, but the inner tag priority must be the same as the outer tag priority in the same MVLAN and cannot be configured separately. If interconnection with a device of non-0x8100 TPID is required, the MVLAN is configured to a stacking VLAN and then TPIDs of inner tag and outer tag are configured globally.

 Copyright © Huawei Technologies Co., Ltd.

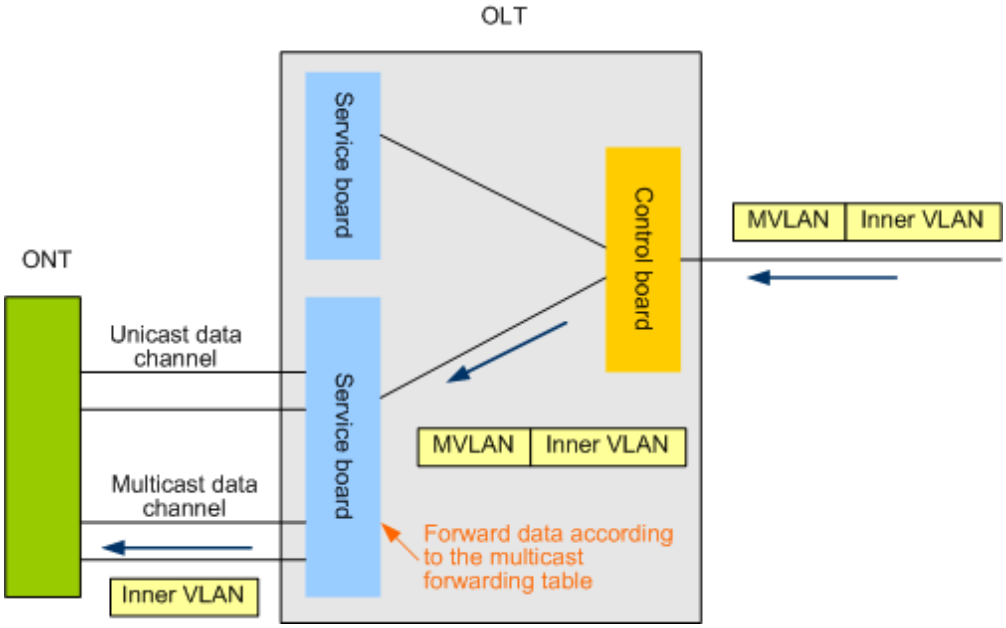
- Multicast data hardware forwarding

  The data forwarding flow of double-tag multicast streams varies according to the GPON duplication mode. If the multicast duplication mode is multicast (single-copy duplication), multicast streams are forwarded on the control board and service board according to the multicast forwarding entry. The VLAN carried by the multicast streams forwarded to the ONT is the CVLAN. Because of the multicast duplication feature, in this scenario, the inner VLAN is often directly defined as the MVLAN (In addition, the device currently supports only this scenario).

  NOTE

  This scenario is applicable only when the TPID is 0x8100.

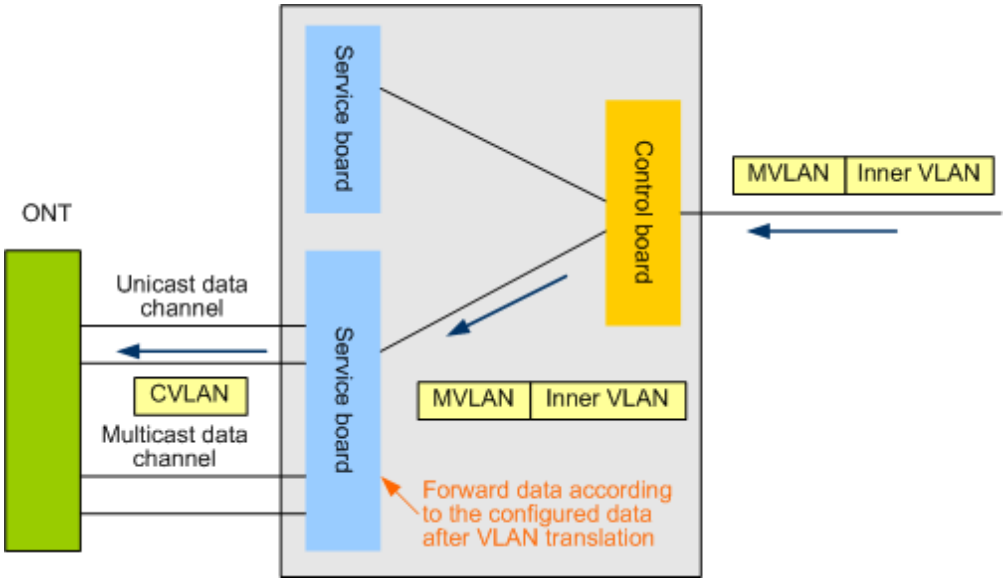
  Figure 13-57 Double-tag multicast hardware forwarding (single-copy duplication)



  If the multicast duplication mode is unicast (multi-copy duplication), multicast streams are forwarded on the control board according to the multicast forwarding entry and on the service board according to the configured service port. The VLAN carried by the multicast streams forwarded to the ONT is the CVLAN. In this scenario, the inner VLAN is often defined as the CVLAN.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-58 Double-tag multicast hardware forwarding (multi-copy duplication)



13.4.5.9 Transparent Transmission of Multicast Data

- Transparent transmission of multicast data for private-line users
  - Transparent transmission of IGMP messages

    In the upstream direction, the service board determines whether to transparently transmit the received IGMP messages according to the VLAN-level IGMP transparent transmission policy and service-port-level transparent transmission policy. If transparent transmission is allowed, IGMP messages are added with the SVLAN tag on the service board and then are transmitted to the control board. After arriving at the control board, IGMP messages are broadcast by the control board within the SVLAN. If transparent transmission is not allowed, IGMP messages are dropped.

    In the downstream direction, the IGMP messages transmitted from the network side are broadcast by the control board within the SVLAN and after they arrive at the service board, the service board determines whether to transparently transmit them according to the VLAN-level IGMP transparent transmission policy and service-port-level transparent transmission policy. If transparent transmission is allowed, the service board translates the SVLAN tag to the CLAN tag according to the configuration of the traffic stream and then transmits the messages to users. If transparent transmission is not allowed, IGMP messages are dropped.
  - Transparent transmission of unknown multicast data

    In the upstream direction, the service board determines whether to transparently transmit the received unknown multicast data according to the VLAN-level IGMP transparent transmission policy and service-port-level transparent transmission policy. If transparent transmission is allowed, IGMP messages are added with the SVLAN tag on the service

 Copyright © Huawei Technologies Co., Ltd.

board and then are transmitted to the control board. After arriving at the control board, IGMP messages are broadcast by the control board within the SVLAN. If transparent transmission is not allowed, IGMP messages are dropped.

In the downstream direction, the unknown multicast data transmitted from the network side is broadcast by the control board within the SVLAN and after the data arrives at the service board, the service board determines whether to transparently transmit the data according to the VLAN-level IGMP transparent transmission policy and service-port-level transparent transmission policy. If transparent transmission is allowed, the service board translates the SVLAN tag to the CLAN tag according to the configuration of the traffic stream and then transmits the data to users. If transparent transmission is not allowed, unknown multicast data is dropped.

NOTE

To prevent the multicast data of the multicast user provisioned with multicast service from being transmitted to the upstream unauthorized multicast sources, make sure that the policy of transmitting unknown multicast data is set to drop. The transparent transmission policies of unknown multicast traffic have the switches of two levels on a service board: the VLAN level and the service port level. When the two switches are both set to transparent transmission, the policy is transparent transmission. When either of the two switches is set to drop, the policy is drop. (Only transparent transmission is supported for connection-oriented traffic and the policy is not configurable in this case.)

- Co-existence of IPTV service and transparent transmission of multicast data
  - Multi-service-port solution

    IPTV service and multicast transparent transmission service are carried on two service ports, and the SVLAN of the service port that carries multicast transparent transmission service must not be the MVLAN.

    The service port that carries IPTV service processes the received IGMP messages following the flow of processing IPTV service, and forwards the multicast data according to the multicast forwarding entry.

    The service port that carries multicast transparent transmission service transparently transmits or drops the received IGMP messages according to the IGMP transparent transmission policy of the traffic stream, and transmits or drops the received unknown multicast data according to the unknown multicast transparent transmission policy of the traffic stream.
  - Single-service-port solution

    IPTV service and multicast transparent transmission service are carried on one service port, whose SVLAN must not be the MVLAN.

    When the Access Node receives upstream IGMP messages, it matches the multicast group address in IGMP messages to the programs in the MVLAN. If the group address successfully matches a program, the Access Node processes the messages as IPTV service. If the group address fails to match any program, the Access Node determines whether to transparently transmit the messages according to the IGMP transparent transmission policy of the SVLAN and service port. The Access Node transparently transmits the messages only when the IGMP transparent transmission policy is enabled for both the SVLAN and service port.

    If the Access Node receives downstream IGMP messages that carry the MVLAN tag, the Access Node processes the IGMP messages as IPTV

 Copyright © Huawei Technologies Co., Ltd.

service. If the messages carry the SVLAN tag, the Access Node forwards them according to the IGMP transparent transmission policy of the SVLAN and service port. The Access Node transparently transmits the messages only when the IGMP transparent transmission policy is enabled for both the SVLAN and service port.

If the Access Node receives the multicast data of IPTV service, the Access Node forwards the multicast data according to the multicast forwarding entry. If the multicast data is unknown, the Access Node forwards the data according to the unknown multicast transparent transmission policy of the VLAN and service port. The Access Node transparently transmits the data only when the unknown multicast transparent transmission policy is enabled for both the SVLAN and service port.

13.5 IPv6 Multicast

This topic describes the aspects unique to IPv6 multicast and the differences between IPv4 multicast and IPv6 multicast.

13.5.1 Introduction to IPv6 Multicast

Overview

As a substitution of IPv4, IPv6 uses a 128-bit address structure to resolve IP address shortage issues and also optimizes some features. The difference between IPv6-based multicast and IPv4-based multicast lies in the significant increase of IP addresses. Other functions of the two types of multicast, such as group member management, multicast packet forwarding, and multicast routing setup, are basically the same.

Purpose

After an IPv4 network evolves to an IPv6 network, IPv6 multicast technologies can be used to provide carriers with a comprehensive set of IPv6 video services, such as live TV and near video on demand (NVoD).

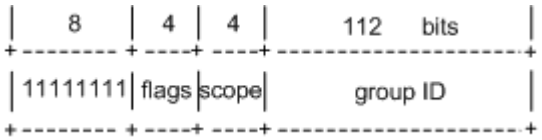
Using IPv6 multicast technologies, the network device can manage, control, and forward IPv6 video services and in this way meets carriers' requirements for provisioning IPv6 video services.

13.5.2 Principle

IPv6 Multicast Address

Figure 13-59 shows the format of an IPv6 multicast address as defined in RFC4291.

Figure 13-59 Format of an IPv6 multicast address



 Copyright © Huawei Technologies Co., Ltd.

The most significant eight bits of an IPv6 multicast address are consistently 11111111; therefore, an IPv6 multicast address always starts with FF.

Flags is a 4-bit field that indicates four flags set on the multicast address. The four flags are defined as |0|R|P|T|. The highest-order flag is always set to 0. The meanings of the other flags are as follows:

Table 13-25 Meaning of the flags Field

Name of the Bit	Value	Description
R	0	Indicates a multicast address that does not embed the unicast address on the rendezvous point (RP) of the multicast group.
	1	Indicates a multicast address that embeds the unicast address on the RP of the multicast group.
P	0	Indicates a multicast address that is not assigned based on the unicast prefix.
	1	Indicates a unicast-prefix-based multicast address.
T	0	Indicates a permanently assigned multicast address.
	1	Indicates a transient (not permanently assigned) multicast address. **NOTE** Currently, the multicast services of the device all use transient multicast addresses. Therefore, the T bit must be set to 1.

Scope is a 4-bit field that indicates the multicast scope. This field is used to signify the scope of a multicast group, that is, whether the multicast group contains nodes of the same local network, site, or organization, or nodes from the global address space. The values of this field are defined as follows:

Table 13-26 Meaning of the Scope Field

Value of the Scope Field	Description
0	Reserved

 Copyright © Huawei Technologies Co., Ltd.

Value of the Scope Field	Description
1	Node/Interface-local scope
2	Link-local scope
3	Reserved
4	Admin-local scope
5	Site-local scope
8	Organization-local scope
E	Global scope
F	Reserved
6, 7, 9, and A-D	Unassigned

Group ID identifies the multicast group.

Ethernet Multicast MAC Address

RFC2464 defines a set of rules for mapping IPv6 multicast addresses to MAC addresses. An IPv6 address is mapped to the MAC address 3333.XXXX.XXXX, with the 32-bit XXXX.XXXX copied from the least significant 32 bits of the IPv6 address.

Protocol Interoperation

IPv6-based group members are managed using Multicast Listener Discovery (MLD).

MLD is a sub-protocol of Internet Control Message Protocol version 6 (ICMPv6). MLD establishes and maintains the multicast group membership between a user host and its immediately neighboring multicast router. MLD can be regarded as the Internet Group Management Protocol (IGMP) in IPv6, as MLD and IPv6 IGMP have similar implementations.

MLD has two versions: MLDv1 and MLDv2. MLDv2 is fully compatible with MLDv1 and includes all basic concepts of MLDv1.

- MLDv1 (defined in RFC2710)

  MLDv1 is derived from IGMPv2 and supports any-source multicast (ASM), but supports source-specific multicast (SSM) through SSM mapping.
- MLDv2 (defined in RFC3810)

  MLDv2 is a translation of IGMPv3 for IPv6 semantics and supports ASM and SSM. Compared with MLDv1 (defined in RFC2710), MLDv2 has the following improvements:
  - Batch report. The destination IP address in the IPv6 header of a report message is always filled in as FF02:0:0:0:0:0:0:16. In addition, the MLD payload can carry multiple group records, reducing the number of report messages between devices. In an MLDv1 report message, the destination

 Copyright © Huawei Technologies Co., Ltd.

IP address must be filled in as the corresponding group IP address, so one MLDv1 report message cannot carry multiple group records.

- Longer maximum response time for a query message. The maximum response time for a query message is extended from 65.5s (in MLDv1) to 8387.5s (5000s supported by Huawei devices). Therefore, MLDv2 is applicable to large-scale networks.
- Source filter. With the source filter function, the host can receive or not receive the multicast data carrying the IP address of a specified multicast source. This function enables the device to better implement source-specific multicast (SSM) and support the multi-ISP scenario. MLDv1 supports only any-source multicast (ASM). The implementation of source filter in IPv6 is similar to that in IPv4.

Charging Mode

For multicast services, carries or Internet service providers (ISPs) usually use two charging modes:

- Fixed charging. In this mode, programs are divided into different packages. Users need to pay a fixed amount of fee for each package in a fixed period (such as on a year or month basis). This charging mode does not restrict multicast users in the program-ordering count or the volume of traffic ordered.
- Pay per view (PPV): In this mode, users are charged according to how many times they order programs.

IPv6 multicast supports rights profiles and rights control and therefore supports fixed charging. IPv6 multicast does not support PPV because it does not support call detail record (CDR) reporting.

Program Ordering Behavior Analysis

IPv6 multicast supports local log recording and does not support log reporting. Log information can be obtained by local queries for analyzing program ordering behavior.

13.6 Configuring the Multicast Service

The MA5800 supports multicast cascading for reducing the number of ports used on the convergence device, and also supports Multiple Spanning Tree Protocol (MSTP) network protection. With these two functions, the network structure is optimized and the multicast service reliability is improved.

13.6.1 Differences Between IPv4 and IPv6 Multicast Configurations

IPv6 multicast refers to the multicast service implemented using the IPv6 protocol. Differences between IPv4 and IPv6 multicast configurations mainly relate to commands and functions. Before configuring IPv6 multicast, it is recommended that you familiarize yourself with the procedures and principles of configuring IPv4 multicast and be aware of the differences between IPv4 and IPv6 multicast configurations.

 Copyright © Huawei Technologies Co., Ltd.

Differences Regarding Commands

NOTE

- The hyphen (-) indicates that the IPv4 command does not have a counterpart in IPv6. In other words, the corresponding multicast feature is not supported in IPv6.
- The multicast commands that are not listed in the table above are the commands supported by both IPv4 and IPv6.

Table 13-27 Differences between IPv4 and IPv6 multicast commands

IPv4	IPv6
Multicast additional functions	
igmp leave-proxy	igmp ipv6 leave-proxy
igmp priority	igmp ipv6 priority
igmp report-proxy	igmp ipv6 report-proxy
igmp echo	-
igmp encapsulation	-
igmp policy	-
igmp query-offline-user	-
igmp user-action-report	-
igmp mismatch	-
display igmp policy	-
Protocol parameters	
igmp proxy router gen-query-interval	igmp ipv6 router gen-query-interval
igmp proxy router gen-response-time	igmp ipv6 router gen-response-time
igmp proxy router robustness	igmp ipv6 router robustness
igmp proxy router sp-query-interval	igmp ipv6 router sp-query-interval
igmp proxy router sp-query-number	igmp ipv6 router sp-query-number
igmp proxy router sp-response-time	igmp ipv6 router sp-response-time
igmp initial-unsolicited-report interval	igmp ipv6 initial-unsolicited-report interval
igmp unsolicited-report interval	igmp ipv6 unsolicited-report interval
display igmp config vlan	display igmp ipv6 config vlan

 Copyright © Huawei Technologies Co., Ltd.

IPv4	IPv6
igmp proxy router timeout	-
Multicast VLAN (MVLAN)	
igmp mode	igmp ipv6 mode
igmp match mode	igmp ipv6 match mode
igmp version	igmp ipv6 version
igmp match group	igmp ipv6 match group
display igmp config vlan	display igmp ipv6 config vlan
igmp send global-leave	-
Multicast users and rights	
debugging igmp	debugging igmp ipv6
Statistics measurement	
igmp statistic reset	igmp ipv6 statistic reset
display igmp statistic	display igmp ipv6 statistic
Logs	
display igmp log statistic	-
display igmp log	-

Differences Regarding Multicast Basic Service Configurations

IPv4 multicast can share VLANs with IPv6 multicast. Therefore, you can deploy IPv6 multicast in existing IPv4 MVLANs just by enabling IPv6 multicast in the IPv4 MVLANs and adding IPv6 multicast programs to the rights profiles. The parameters that have already been configured for users, such as the bound rights profiles and MVLANs, remain unchanged.

13.6.2 Configuring the Multicast Service on a Single NE

When the network structure is simple, the configuration of a single NE can meet multicast service requirements. Compared with cascading networking, single-NE networking is more secure and stable, and provides more bandwidth resources, but requires more line resources. The method of configuring multicast services for an NE in the cascading or MSTP networking scenario is the same as that in single-NE networking scenario.

Application Context

The multicast feature of the MA5800 is mainly applied to the live TV and near-video on demand (NVOD) multicast video services. The MA5800 runs the IGMP

 Copyright © Huawei Technologies Co., Ltd.

proxy or IGMP snooping protocol, and the interconnected device can run the IGMP proxy, IGMP snooping, or multicast router protocol.

Currently, the multicast application of the MA5800 is oriented to Layer 2, and the MA5800 forwards data based on VLAN ID+multicast MAC address. A multicast program in the network is identified by VLAN ID + multicast IP address uniquely. The MA5800 differentiates multicast sources by VLAN ID. It allocates a unique VLAN ID to each multicast source, controls the multicast domain and the user right based on the multicast VLAN ID, and provides a platform for different ISPs to implement different multicast video services.

Data preparation

Before configuring the multicast video service, plan the data items as listed in **Table 13-28**.

Table 13-28 Data plan for configuring the multicast service on a single MA5800

Device	Data Item
MA5800	Multicast VLAN
	Layer 2 multicast protocol
	IGMP version of the multicast VLAN
	IGMP version of the multicast user
	Multicast program configuration mode
	Multicast general query and group-specific query parameters **NOTE** The default values are adopted.
	Program list
	User authentication policy
	Program bandwidth, upstream port bandwidth, and user bandwidth
	Multicast ONT
	Multicast logging policy
Upper-layer multicast router	IGMP version **NOTE** The IGMP version of the upper-layer multicast router must not be earlier than the IGMP version of the multicast VLAN used by the MA5800.

 Copyright © Huawei Technologies Co., Ltd.

Device	Data Item
Home gateway or modem	IGMP version **NOTE** The IGMP version of the home gateway or modem must not be earlier than the IGMP version of the multicast user on the MA5800.

Table 13-29 Default settings of the multicast service

Feature	Default Settings
Multicast protocol	Disabled
IGMP version	V3
Multicast program configuration mode	Static configuration mode
Multicast bandwidth management	Enabled
Multicast preview	Enabled
Multicast log function	Enabled

Configuration Flowchart

13.6.2.1 Configuring Multicast Global Parameters

The general parameters of Layer 2 multicast protocols (including IGMP proxy and IGMP snooping) configured for a device are applicable to all the multicast VLANs on the device.

Context

The multicast global parameters include general query, group-specific query, the policy of processing multicast packets and the multicast forwarding mode.

The description of a general query is as follows:

- Purpose: A general query packet is periodically sent by the MA5800 to check whether there is any multicast user who leaves the multicast group without sending the leave packet. Based on the query result, the MA5800 periodically updates the multicast forwarding table and releases the bandwidth of the multicast user that has left the multicast group.
- Principle: The MA5800 periodically sends the general query packet to all online IGMP users. If the MA5800 does not receive the response packet from a multicast user within a specified time (Robustness variable x General query interval + Maximum response time of a general query), it regards the user as having left the multicast group and deletes the user from the multicast group.

The description of a group-specific query is as follows:

 Copyright © Huawei Technologies Co., Ltd.

- Purpose: A group-specific query packet is sent by the MA5800 after a multicast user that is not configured with the leave attribute sends the leave packet. The group-specific query packet is used to check whether the multicast user has left the multicast group.
- Principle: When a multicast user leaves a multicast group, for example, switches to another channel, the user unsolicitedly sends a leave packet to the MA5800. If the multicast user is not configured with the leave attribute, the MA5800 sends a group-specific query packet to the multicast group. If the MA5800 does not receive the response packet from the multicast user within a specified duration (Robustness variable x Group-specific query interval + Maximum response time of a group-specific query), it deletes the multicast user from the multicast group.

NOTE

The configuration steps for IPv4 multicast and IPv6 multicast are similar but detailed commands are different. This topic describes the configuration steps for IPv4 multicast, and provides the configuration example for IPv6 multicast.

Table 13-30 lists the default settings of the multicast global parameters. In the actual application, you can modify the values according to the data plan.

Table 13-30 Default settings of the multicast global parameters

Parameter	Default Value
General query parameter	Query interval: 125s Maximum response time: 10s Robustness variable (query times): 2
Group-specific query parameter	Query interval: 1s Maximum response time: 0.8s. Robustness variable (query times): 2
Policy of processing multicast packets	IGMP packet: normal (IGMP packets are processed as controllable multicast) Unknown multicast packet: • For switch-oriented traffic streams: discard • For connection-oriented traffic streams: transparent transmission

Procedure

Step 1 Configure the general query parameters.

1. Run the **igmp proxy router gen-query-interval** command to set the general query interval. By default, the general query interval is 125s.
2. Run the **igmp proxy router gen-response-time** command to set the maximum response time of the general query. By default, the maximum response time of the general query is 10s.

 Copyright © Huawei Technologies Co., Ltd.

3. Run the **igmp proxy router robustness** command to set the robustness variable (query times) of the general query. By default, the robustness variable (query times) is 2.

Step 2 Set the group-specific query parameters.

1. Run the **igmp proxy router sp-query-interval** command to set the group-specific query interval. By default, the group-specific query interval is 1s.
2. Run the **igmp proxy router sp-response-time** command to set the maximum response time of the group-specific query. By default, the maximum response time of the group-specific query is 0.8s.
3. Run the **igmp proxy router sp-query-number** command to set the robustness variable (query times) of the group-specific query. By default, the robustness variable (query times) is 2.

Step 3 Configure the policy of processing multicast packets.

The default values are adopted for multicast service and do not need to be modified. To control the forwarding of multicast packets when configuring other services, run the following commands to configure the policy.

1. Run the **igmp policy** command to set the policy of processing IGMP packets.
2. Run the **multicast-unknown policy** command to set the policy of processing unknown multicast packets(downstream UDP packets).

Step 4 Run the **display igmp config global** command to check whether the values of the multicast parameters are correct.

----End

Example

Example: To set the query interval to 150s, maximum response time to 20s, and number of queries to 3 in the multicast general query parameters for the IPv4 multicast VLAN 100 in IGMP V3, do as follows:

```
huawei(config)#multicast-vlan 100
huawei(config-mvlan100)#igmp proxy router gen-query-interval 150
huawei(config-mvlan100)#igmp proxy router gen-response-time v3 200
huawei(config-mvlan100)#igmp proxy router robustness 3
```

Example: To set the query interval to 20s, maximum response time to 10s, and number of queries to 3 in the multicast group-specific query parameters for the IPv4 multicast VLAN 100 in IGMP V3, do as follows:

```
huawei(config)#multicast-vlan 100
huawei(config-mvlan100)#igmp proxy router sp-query-interval 200
huawei(config-mvlan100)#igmp proxy router sp-response-time v3 100
huawei(config-mvlan100)#igmp proxy router sp-query-number 3
```

Example: To set the query interval to 150s, maximum response time to 20s, and number of queries to 3 in the multicast general query parameters for the IPv6 multicast VLAN 200 in MLD V2, do as follows:

```
huawei(config)#multicast-vlan 200
huawei(config-mvlan200)#igmp ipv6 router gen-query-interval 150
huawei(config-mvlan200)#igmp ipv6 router gen-response-time v2 200
huawei(config-mvlan200)#igmp ipv6 router robustness 3
```

 Copyright © Huawei Technologies Co., Ltd.

Example: To set the query interval to 20s, maximum response time to 10s, and number of queries to 3 in the multicast group-specific query parameters for the IPv6 multicast VLAN 200 in MLD V2, do as follows:

```
huawei(config)#multicast-vlan 200
huawei(config-mvlan200)#igmp ipv6 router sp-query-interval 200
huawei(config-mvlan200)#igmp ipv6 router sp-response-time v2 100
huawei(config-mvlan200)#igmp ipv6 router sp-query-number 3
```

13.6.2.2 Configuring the Multicast VLAN and the Multicast Program

In the application of multicast service, multicast VLANs (MVLANs) are used to distinguish multicast ISPs. Generally, an MVLAN is allocated to each multicast ISP for the VLAN-based management of multicast programs, multicast protocols, IGMP versions, and the VLAN-based control of multicast domain and user right.

Context

To create a multicast VLAN, a common VLAN must be created first. The multicast VLAN can be the same as the unicast VLAN. In this case, the two VLANs can share the same service stream channel. The multicast VLAN can be different from the unicast VLAN. In this case, the two VLANs use different service stream channels.

One user port can be added to multiple multicast VLANs under the following restrictions:

- Among all the multicast VLANs of a user port, only one multicast VLAN is allowed to have dynamically generated programs.
- One user port is not allowed to belong to multiple MVLANs that are in the IGMPv3 snooping mode.

The source IP address in the multicast packets that are sent to the upper device by the OLT may be as follows:

- If the IP address of the program VLAN interface is configured, the source IP address is the IP address of VLAN interface.
- If the IP address of the program VLAN interface is not configured, the source IP address is the host IP address of the program.
- If the host IP address is not configured, the default address 0.0.0.0 is used.

Table 13-31 lists the default settings of the MVLAN attributes, including the Layer 2 multicast protocol, IGMP version, multicast program, and multicast upstream port.

Table 13-31 Default settings of the MVLAN attributes

Parameter	Default Value
Program matching mode	enable (static configuration mode)
Multicast upstream port mode	default
Layer 2 multicast protocol	off (multicast function disabled)
IGMP version	v3

 Copyright © Huawei Technologies Co., Ltd.

Parameter	Default Value
Priority of forwarding IGMP packets by the upstream port	6
Group filter mode	asm-ssm

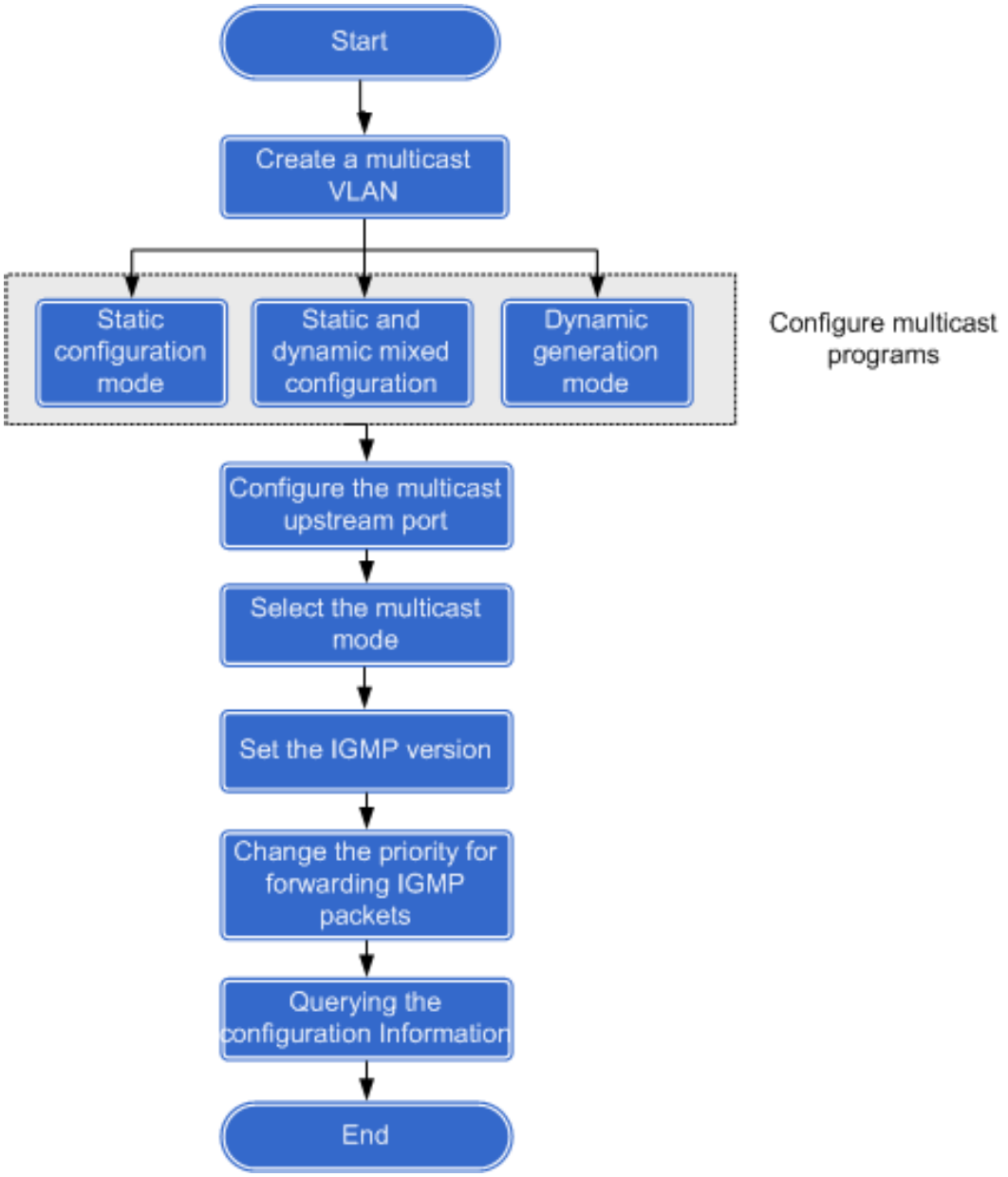
NOTE

The device supports IPv4 and IPv6 multicast services. The two types of services have the same VLAN configurations and only differ in commands. The procedure in this topic uses IPv4 multicast as an example, and the example uses IPv6 multicast as an example.

Configuration Flowchart

Figure 13-60 shows the flowchart for configuring the multicast VLAN and multicast program.

Figure 13-60 Flowchart for configuring the multicast VLAN and multicast program



Procedure

Step 1 Create a multicast VLAN.

1. Run the **vlan** command to create a VLAN, and set the VLAN type according to the actual application. For details on the VLAN configuration, see **Configuring VLAN**.
2. Run the **multicast-vlan** command to set the created VLAN to a multicast VLAN. The VLAN with S+C forwarding mode cannot be set as a multicast VLAN.

Step 2 Configure multicast programs.

 Copyright © Huawei Technologies Co., Ltd.

The multicast VLAN can be configured statically or generated dynamically. The program configuration of the MVLAN has three modes: static configuration, dynamic generation, and static and dynamic mixed configuration.

- Static configuration mode: Configure the program list before users watch the video programs. In this mode, the right profile can be used to implement controllable multicast. The program list and the right profile, however, need to be maintained according to the change of the video service. The program host, program prejoin, and multicast bandwidth management functions are supported.

1. Run the **igmp match mode enable** or **igmp ipv6 match mode enable** command to set the static configuration mode. By default, the system adopts the static configuration mode.
2. Run the **igmp program add** command to add a multicast program.

   NOTE

   If the IGMP version of a multicast VLAN is v3, the program must be configured with a source IP address. If the IGMP version of a multicast VLAN is v2, the program must not be configured with a source IP address.
3. Add a right profile.

   In the BTV mode, run the **igmp profile add** command to add a right profile.
4. Bind the program to the right profile.

   In the BTV mode, run the **igmp profile** command to bind the program to the right profile, and set the right to watch.

   NOTE

   When a user is bound to multiple right profiles, and the right profiles have different rights to a program, the right with the highest priority prevails. You can run the **igmp right-priority** command to adjust the priorities of the four rights: watch, preview, forbidden, and idle. By default, the priorities of the four rights are forbidden > preview > watch > idle.

- Dynamic generation mode: A program list is dynamically generated according to the programs requested by users. In this mode, the program list does not need to be configured or maintained; however, the functions such as program management, user multicast bandwidth management, program preview, and program prejoin are not supported.

1. Run the **igmp match mode disable** or **igmp ipv6 match mode disable** command to set the dynamic generation mode.

   NOTICE

   The **igmp match mode** or**igmp ipv6 match mode** command can be executed only when the multicast mode is disabled. If the multicast mode is set to off, users are forced to go offline. You are advised to perform initial configuration before service provisioning.
2. Run the **igmp match group** or **igmp ipv6 match group** command to configure the IP address range of the program group that can be dynamically generated. Users can order only the programs whose IP addresses are within the specified range.

 Copyright © Huawei Technologies Co., Ltd.

- Static and dynamic mixed configuration: Configure the program list before users watch the video programs. In this mode, the right profile can be used to implement controllable multicast. The program list and the right profile, however, need to be maintained according to the change of the video service. The program host, program prejoin, and multicast bandwidth management functions are supported.

1. Run the **igmp match mode disable** or **igmp ipv6 match mode disable** command to set the mode to the dynamic generation mode.
2. Run the **igmp match group** or **igmp ipv6 match group** command to configure the IP address range of the program group that can be dynamically generated. Users can order only the programs whose IP addresses are within the specified range.
3. Run the **igmp program add [name name] ip ip-addr [sourceip ip-addr] [hostip ip-addr]** command to add a multicast static program.

   NOTE

   When the range of static program IP addresses and the range of dynamic program IP addresses overlap each other, static programs can go online with priority.
4. Run the **igmp group-filter-mode** command to set the group filter mode based on multicast VLAN (MVLAN).

   NOTE

   – When the group filter mode of an MVLAN is configured to **asm-only** or **asm-ssm**, only one program with the unique multicast IP address is generated in the MVLAN. The [*, G] multicast forwarding table is used for this MVLAN instance on the forwarding plane.
   – When the group filter mode of an MLVAN is configured to **ssm-only**, multiple programs with the same multicast IP addresses but different source IP addresses can be generated in the MVLAN. The [s, g] multicast forwarding table is used for this VLAN instance on the forwarding plane.

     NOTE

     The source IP addresses are regarded as different ones when they have different least significant 20 bits from each other.
   – The maximum number of programs is calculated according to the number of actually-generated programs. For example:

     ■ When a multicast user joins an MVLAN with the multicast filter mode **asm-ssm** and the system receives two packets with IP addresses [S1, G1] and [S2, G1], the system generates only one multicast program with the multicast IP address G1 for the multicast user;

     ■ When a multicast user joins an MVLAN with the multicast filter mode **ssm-only** and the system receives two packets with IP addresses [S1, G1] and [S2, G1], the system generates two multicast programs with IP addresses [S1, G1] and [S2, G1].

Step 3 Configure the multicast upstream port.

1. Run the **igmp uplink-port** command to configure the multicast upstream port. The packets of the MVLAN corresponding to the upstream port are forwarded and received by this upstream port.
2. In the BTV mode, run the **igmp uplink-port-mode** command to change the mode of the multicast upstream port. By default, the port is in the default mode. In the MSTP network, the port adopts the MSTP mode.

 Copyright © Huawei Technologies Co., Ltd.

  - Default mode: If the MVLAN contains only one upstream port, the IGMP packets that go upstream can be sent only by this port. If the MVLAN contains multiple upstream ports, the IGMP packets that go upstream are sent by all the upstream ports.
  - MSTP mode: This mode is adopted in the MSTP network.

Step 4 Select the multicast mode.

Run the **igmp mode { proxy | snooping }** command to select the Layer 2 multicast mode. By default, the multicast mode is disabled.

In terms of multicast processing mode, the MA5800 supports the Internet Group Management Protocol (IGMP) Proxy and IGMP Snooping Layer 2 multicast protocols. IGMP proxy and IGMP snooping both support multicast video data forwarding; however, the two modes have different processing mechanisms.

- In IGMP snooping, the related information for maintaining multicast forwarding entries is obtained by listening to the IGMP packets between the user and the multicast router.
- IGMP proxy intercepts the IGMP packets between the user and the multicast router, processes the IGMP packets, and then forwards the IGMP packets to the upper-layer multicast router. For the multicast user, the MA5800 is a multicast router that implements the router functions in the IGMP protocol; for the multicast router, the MA5800 is a multicast user.

In the IGMP snooping mode, proxy can be enabled for the report packet and the leave packet. When a multicast user joins or leaves a multicast program, the MA5800 can implement IGMP proxy. IGMP snooping and IGMP proxy are controlled separately.

- Run the **igmp report-proxy enable** command to enable the proxy of the snooping report packet. When the first user requests to join a program, after authenticating the user, the MA5800 sends the user report packet to the network side and receives a corresponding multicast stream from the multicast router. The report packets of the users who follow the first user are not sent by the MA5800 to the network side.
- Run the **igmp leave-proxy enable** command to enable the proxy of the snooping leave packet. When the last user requests to leave the program, the MA5800 sends the user leave packet to the network side to request the upper-layer device to stop sending multicast streams. The leave packets of the users who precede the last user are not sent by the MA5800 to the network side.

Step 5 Set the IGMP version.

Run the **igmp version{ v2 | v3 }** command to set the IGMP version. By default, IGMP v3 is enabled in the system. If the upper-layer and lower-layer devices in the network are IGMP v2 devices and cannot recognize the IGMP v3 packets, run this command to change the IGMP version.

IGMP v3 is compatible with IGMP v2 in packet processing. If IGMP v3 is enabled on the MA5800 and the upper-layer multicast router switches to IGMP v2, the MA5800 automatically switches to IGMP v2 when receiving the IGMP v2 packets. If the MA5800 does not receive any more IGMP v2 packets within the preset IGMP v2 timeout time, it automatically switches back to IGMP v3. In the BTV mode, run

 Copyright © Huawei Technologies Co., Ltd.

the **igmp proxy router timeout** command to set the IGMP v2 timeout time. By default, the timeout time is 400s.

Step 6 Change the priority for forwarding IGMP packets.

Run the **igmp priority** command to change the priority for forwarding the IGMP packets by the upstream port. By default, the priority is 6 and does not need to be changed.

- In the IGMP proxy mode, the IGMP packets sent from the upstream port to the network side adopt the priority set through the preceding command in the MVLAN.
- In the IGMP snooping mode, the IGMP packets forwarded to the network side adopt the priority of the user service stream. The priority of the service stream is set through the traffic profile.

Step 7 Check whether the configuration is correct.

- Run the **display igmp config vlan** command to query the attributes of the multicast VLAN.
- Run the **display igmp program vlan** command to query the information about the program of the MVLAN.

----End

Example

To configure the MVLAN and multicast program for the IPv4 multicast, do as follows:

- MVLAN ID: 101
- Program configuration mode: static configuration; program IP address: 224.1.1.1
- Source IP address: 10.10.10.10; host IP address: 10.0.0.254
- Program bandwidth: 5000 kbit/s
- MVLAN upstream port: 0/9/0
- Protocol: IGMP proxy; IGMP version: v3
- Group filter mode: ssm-only

```
huawei(config)#vlan 101 smart
huawei(config)#multicast-vlan 101
huawei(config-mvlan101)#igmp match mode enable
huawei(config-mvlan101)#igmp program add name movie ip 224.1.1.1 sourceip 10.10.10.10
hostip 10.0.0.254 bandwidth 5000
huawei(config-mvlan101)#igmp uplink-port 0/9/0
huawei(config-mvlan101)#igmp mode proxy
  Are you sure to change IGMP mode?(y/n)[n]:y
  Command is being executed. Please wait...
  Command has been executed successfully
huawei(config-mvlan101)#igmp version v3
huawei(config-mvlan101)#igmp group-filter-mode ssm-only
```

To configure the MVLAN and multicast program for the IPv4 multicast, do as follows:

- MVLAN ID: 101
- Program configuration mode: dynamic generation

 Copyright © Huawei Technologies Co., Ltd.

- Address range of the dynamic program group: 224.1.1.10 to 224.1.1.50
- MVLAN upstream port: 0/9/0
- Protocol: IGMP proxy; IGMP version: v3

```
huawei(config)#vlan 101 smart
huawei(config)#multicast-vlan 101
huawei(config-mvlan101)#igmp match mode disable
  This operation will delete all the programs in current multicast vlan
  Are you sure to change current match mode? (y/n)[n]: y
  Command is being executed, please wait...
  Command has been executed successfully
huawei(config-mvlan101)#igmp match group ip 224.1.1.10 to-ip 224.1.1.50
huawei(config-mvlan101)#igmp uplink-port 0/9/0
huawei(config-mvlan101)#igmp mode proxy
  Are you sure to change IGMP mode?(y/n)[n]:y
  Command is being executed. Please wait...
  Command has been executed successfully
huawei(config-mvlan101)#igmp version v3
```

To configure the MVLAN and multicast program for the IPv4 multicast, do as follows:

- MVLAN ID: 101
- Program configuration mode: static and dynamic mixed configuration
- MVLAN upstream port: 0/9/0
- IP address of the static program: 224.1.1.1; source IP address: 10.10.10.10; host IP address: 10.0.0.254; program bandwidth: 5000 kbit/s
- Address range of the dynamic program group: 224.1.1.10 to 224.1.1.50
- Protocol: IGMP proxy; IGMP version: v3

```
huawei(config)#vlan 101 smart
huawei(config)#multicast-vlan 101
huawei(config-mvlan101)#igmp match mode disable
  This operation will delete all the programs in current multicast vlan
  Are you sure to change current match mode? (y/n)[n]: y
  Command is being executed, please wait...
  Command has been executed successfully
huawei(config-mvlan101)#igmp match group ip 224.1.1.10 to-ip 224.1.1.50
huawei(config-mvlan101)#igmp program add name movie ip 224.1.1.1 sourceip 10.10.10.10
hostip 10.0.0.254 bandwidth 5000
huawei(config-mvlan101)#igmp uplink-port 0/9/0
huawei(config-mvlan101)#igmp mode proxy
  Are you sure to change IGMP mode?(y/n)[n]:y
  Command is being executed. Please wait...
  Command has been executed successfully
huawei(config-mvlan101)#igmp version v3
```

To configure the MVLAN and multicast program for the IPv6 multicast, do as follows:

- MVLAN ID: 101
- Program configuration mode: static configuration; program IPv6 address:ffff::1
- Source IPv6 address: 2001:db8::1
- Program bandwidth: 5000 kbit/s
- MVLAN upstream port: 0/9/0
- Protocol: IGMP proxy; IGMP version: v2

```
huawei(config)#vlan 101 smart
huawei(config)#multicast-vlan 101
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-mvlan101)#igmp ipv6 match mode enable
huawei(config-mvlan101)#igmp program add name movie ipv6 ffff::1 source-ipv6 2001:db8::1
bandwidth 5000
huawei(config-mvlan101)#igmp uplink-port 0/9/0
huawei(config-mvlan101)#igmp ipv6 mode proxy
  Are you sure to change IGMP mode?(y/n)[n]:y
  Command is being executed. Please wait...
  Command has been executed successfully
huawei(config-mvlan101)#igmp ipv6 version v2
```

13.6.2.3 Configuring the Multicast GPON ONT

When the MA5800 is connected with an ONT or an MDU, you need to configure the multicast interconnection data to forward the multicast traffic streams.

Prerequisites

Before configuring a multicast GPON ONT, you must add the ONT correctly. For the configuration method, see **Configuring the GPON ONT**.

Context

- When the OLT is connected with an ONT, the MA5800 manages the ONT in the OMCI mode. In this case, you need to configure the ONT line profile and the ONT service profile, configure the multicast data in the ONT service profile, and bind the profiles to the ONT to issue the multicast service.
- When the OLT is connected with an MDU, the MA5800 manages the MDU in the SNMP mode. In this case, you do not need to configure the ONT service profile. You only need to configure the multicast data on the MDU interconnected with the MA5800 to forward the multicast traffic streams.

Procedure

Step 1 Add an ONT line profile.

For the configuration method, see **Configuring the GPON ONT Line Profile**.

Step 2 Add an ONT service profile.

Run the **ont-srvprofile gpon** command to add a GPON ONT service profile, and then enter the GPON ONT service profile mode.

If the ONT management mode is the SNMP mode, you do not need to configure the service profile. After adding a GPON ONT service profile, directly enter the GPON ONT service profile mode to configure the related multicast data.

1. Run the **ont-port** command to configure the port capability set of the ONT. The port capability set in the ONT service profile must be the same as the actual ONT capability set.
2. Run the **port vlan** command to configure the port VLAN of the ONT.
3. Configure the multicast mode of the ONT.

   Run the **multicast mode { igmp-snooping|olt-control|unconcern }** command to select the multicast mode.

 Copyright © Huawei Technologies Co., Ltd.

- **igmp-snooping**: IGMP snooping obtains the related information and maintains the multicast forwarding entries by listening to the IGMP packets in the communication between the user and the multicast router.
- **olt-control**: It is the dynamic controllable multicast mode. A multicast forwarding entry can be created for the multicast join packet of the user only after the packet passes the authentication.
- **unconcern**: It is the unconcern mode. After this mode is selected, the OLT does not limit the multicast mode, and the multicast mode on the OLT automatically matches the multicast mode on the ONT.

4. (Optional)Configure the multicast forwarding mode.

   Run the **multicast-forward { untag | tag { translation vlanid | transparent } | unconcern** command to configure the multicast forwarding mode and multicast forwarding VLAN. The forwarding mode is not concerned by default.

   - **tag**: Specifies the multicast forwarding mode as tag. If the VLAN tag of the multicast packet needs to be transparently transmitted, use **transparent**; if the VLAN tag of the multicast packet needs to be switched, use **translation** and set the VLAN tag used after the switching. When the ONT is directly connected to the home gateway in the application, use this parameter.
   - **untag**: Specifies the multicast forwarding mode as untag, that is, the downstream multicast packet from the ONT's Ethernet port to a next directly connected device does not carry the VLAN tag. When the ONT is directly connected to the set top box (STB) or PC, use this parameter.
   - **unconcern**: Indicates that the multicast forwarding mode is not concerned. When the ONT multicast mode need not be configured by the OLT and is determined by the ONT condition, use **unconcern**. This value is the default value.

5. After the configuration is complete, run the **commit** command to make the configured service profile take effect.

   NOTE

   For an ONT that is added through the **ont add** command or an automatically found ONT that is confirmed through the **ont confirm** command, if you run the **commit** command after modifying the ONT line profile parameters and the ONT service profile parameters, the modified profile parameters take effect immediately.

----End

Example

To configure the ONT service profile 10 of 4 ETH ports, 2 POTS ports, the VLAN of the ETH port as 10, the multicast mode as IGMP snooping, the multicast forwarding mode as unconcern, do as follows:

```
huawei(config)#ont-srvprofile gpon profile-id 10
huawei(config-gpon-srvprofile-10)#ont-port eth 4 pots 2
huawei(config-gpon-srvprofile-10)#port vlan eth 1 10
huawei(config-gpon-srvprofile-10)#multicast mode igmp-snooping
huawei(config-gpon-srvprofile-10)#multicast-forward unconcern
huawei(config-gpon-srvprofile-10)#commit
huawei(config-gpon-srvprofile-10)#quit
```

 Copyright © Huawei Technologies Co., Ltd.

13.6.2.4 Configuring a Multicast User

This topic describes how to configure a multicast user and the related user right for provisioning the multicast service.

Prerequisites

Before configuring a multicast user, create a service channel. The procedure is as follows:

- Configuring a GPON multicast user
  a. **Configure the VLAN.**
  b. **Configure the upstream port.**
  c. **Configure the multicast GPON ONT.**
  d. **Configure the GPON user port.**
  e. **Configure the GPON traffic stream.**

Context

Add a multicast user, and bind the multicast user to the multicast VLAN to create a multicast member. Bind the multicast user to a right profile to implement multicast user authentication.

Table 13-32 lists the default settings of the attributes related to the multicast user.

Table 13-32 Default settings of the attributes related to the multicast user

Parameter	Default Value
Maximum number of programs that can be watched by the multicast user	8
Maximum number of programs of different priorities that can be watched by the multicast user	no limit
Quick leave mode of the multicast user	MAC-based
Global switch of multicast user authentication	enable
IGMP version of the multicast user	v3

Procedure

Step 1 In the global config mode, run the **btv** command to enter the BTV mode.

Step 2 Configure a multicast user and the multicast user attributes.

1. Add a multicast user.

   Run the **igmp user add service-port** command to add a multicast user.

 Copyright © Huawei Technologies Co., Ltd.

2. Configure the maximum number of programs that can be watched by the multicast user.
   - Run the **igmp user add service-port index max-program { max-program-num | no-limit }** command to configure the maximum number of programs that can be watched by the multicast user concurrently.
   - Run the **igmp user watch-limit service-port { hdtv | sdtv | streaming-video }** command to configure the maximum number of programs of different priorities that can be watched by the multicast user.
3. Set the quick leave mode of the multicast user.

   Run the **igmp user add service-port index quickleave { immediate | disable | mac-based }** command to configure the quick leave mode of the multicast user. By default, the quick leave mode is the MAC-based mode.
   - **Immediate**: After receiving the leave packet of the multicast user, the system immediately deletes the multicast user from the multicast group.
   - **Disable**: After receiving the leave packet of the multicast user, the system sends an ACK packet to confirm that the multicast user leaves, and then deletes the multicast user from the multicast group.
   - **MAC-based**: It is the quick leave mode based on the MAC address. The system checks the MAC address in the leave packet of the user. If it is the same as the MAC address in the report packet of the user and it is the last MAC address of multicast user, the system immediately deletes the multicast user from the multicast group. Otherwise, the system does not delete the multicast user. This mode is applied to the scenario with multiple terminals.
4. Set the IGMP version for the multicast user.

   Run **igmp user add service-port index igmp-version { v2 | v3 | v3-forced | v2-with-query }** command to set the IGMP version for the multicast user. Each multicast user has an independent querier instance. This command specifies the IGMP version (default: v3) for the multicast user querier.
   - v2: specifies the IGMP version to v2 for the multicast user querier. When this setting applies, the system processes only IGMP v2 packets and directly drops IGMP v1 packets and IGMP v3 packets.
   - v3: specifies the v3–compatible mode (default setting for the system). When this setting applies, the system automatically specifies the IGMP version according to the version of the IGMP packets sent by users, but it directly drops IGMP v1 packets.
   - v3-forced: forcibly specifies the IGMP version to V3 for the multicast user querier. When this setting applies, the system processes only IGMP v3 packets but directly drops IGMP v1 packets and IGMP v2 packets.
   - v2-with-query: specifies the user querier version to IGMP v2 and can receive the IGMP v3 packet sent from the user side. When the system receives an IGMP v3 packet sent from the user side, the system sends a common query packet of the IGMP v2 version and discards the IGMP v3 packet.

Step 3 Configure multicast user authentication.

To control the right of a multicast user, you can enable the multicast user authentication function.

 Copyright © Huawei Technologies Co., Ltd.

1. Configure the multicast user authentication function.

   Run the **igmp user add service-port index { auth | no-auth }** command to configure whether to authenticate a multicast user.

   NOTE

   After configuring multicast user authentication, you need to enable the global authentication function to make the configuration take effect. By default, the global authentication function is enabled. You can run the **igmp proxy authorization** command to change the configuration.

2. Bind the multicast user to the right profile. This operation is to implement user authentication.

   Run the **igmp user bind-profile** command to bind the user to a right profile. After the binding, the multicast user has the rights to the programs as configured in the profile.

Step 4 Bind the multicast user to a multicast VLAN.

In the multicast VLAN mode, run the **igmp multicast-vlan member** command to bind the user to the multicast VLAN. Then, the multicast user becomes a multicast member of the multicast VLAN and can request the programs configured in the multicast VLAN.

Step 5 Run the **display igmp user** command to check whether the related multicast user information is correctly configured.

----End

Example

To add multicast user (port) 0/2/1 to multicast VLAN 101, enable user authentication, enable log report, set the maximum bandwidth to 10 Mbit/s, set IGMP version of the multicast user to v3-forced, and bind the user to right profile **music**, do as follows:

```
huawei(config)#service-port 100 vlan 101 gpon 0/2/1 ont 1 eth 0 multi-service user-vlan 101 rx-cttr 2 tx-cttr 2
huawei(config)#btv
huawei(config-btv)#igmp user add service-port 100 auth log enable max-bandwidth 10240 igmp-version v3-forced
huawei(config-btv)#igmp user bind-profile service-port 100 profile-name music
huawei(config-btv)#quit
huawei(config)#multicast-vlan 101
huawei(config-mvlan101)#igmp multicast-vlan member service-port 100
```

13.6.2.5 (Optional) Configuring the Multicast Bandwidth

To limit the multicast bandwidth of a user, you can enable multicast bandwidth management, that is, connection admission control (CAC), and then control the bandwidth of a multicast user by setting the program bandwidth and the user bandwidth.

Prerequisites

The program matching mode of the multicast VLAN must be the static configuration mode.

 Copyright © Huawei Technologies Co., Ltd.

Context

If the CAC function is enabled and a user requests a multicast program, the system compares the remaining bandwidth of the user (bandwidth configured for the user - total bandwidth of the online programs of the user) with the bandwidth of the multicast program. If the remaining bandwidth of the user is sufficient, the system adds the user to the multicast group. If the bandwidth is insufficient, the system does not respond to the request of the user.

If the CAC function is disabled, the system does not guarantee the bandwidth of the multicast program. When the bandwidth is not guaranteed, problems such as mosaic and delay occur in the multicast program.

Table 13-33 lists the default settings of the CAC parameters.

Table 13-33 Default settings of the CAC parameters

Parameter	Default Value
Global CAC function	enable
Bandwidth of the multicast program	5000 kbit/s
Bandwidth of the multicast user	no-limit
Bandwidth of the GPON port	716800 kbit/s

Procedure

Step 1 In the global config mode, run the **btv** command to enter the BTV mode.

Step 2 Enable the global CAC function.

By default, the global CAC function is already enabled. You can run the **igmp bandwidthCAC { enable | disable }** command to change the setting.

Step 3 Configure the bandwidth of the multicast user.

Run the **igmp user add service-port index max-bandwidth** command to allocate the maximum bandwidth of the multicast user.

Step 4 Configure the bandwidth of the multicast program.

- Run the **igmp program add ip ip-addr bandwidth** command to configure the bandwidth of a single multicast program. The program bandwidth is an attribute of a multicast program, specifying the bandwidth requirement of the program being played.
- Run the **igmp bandwidth port frameid/slotid/portid max-bandwidth{ bandwidth | no-limit }** command to configure the program bandwidth of a physical port on a board. This command is available for only the GPON port. The default bandwidth of a port is 716800 kbit/s. Configuring the total program bandwidth for a single port is a way of traffic management, which helps avoid network congestion caused by the excessively-large total program bandwidth on a port. When the total program bandwidth of a port

 Copyright © Huawei Technologies Co., Ltd.

exceeds the value configured using the **igmp bandwidth port frameid/slotid/portid max-bandwidth{ bandwidth | no-limit }** command, subsequent programs ordered by users on this port cannot be played.

Step 5 Check whether the multicast bandwidth configuration is correct.

- Run the **display igmp config global** command to check the status of the global CAC function.
- Run the **display igmp program** command to query the bandwidth allocated to the multicast program.
- Run the **display igmp user** command to query the maximum bandwidth and the occupied bandwidth of the multicast user.

----End

Example

To enable bandwidth management for multicast users, set the user bandwidth to 10 Mbit/s when adding multicast user 0/2/1, and configure the program bandwidth to 1 Mbit/s when adding multicast program 224.1.1.1.

```
huawei(config)#btv
huawei(config-btv)#igmp bandwidthcAC enable
huawei(config-btv)#igmp user add port 0/2/1 max-bandwidth 10240
huawei(config-btv)#quit
huawei(config)#multicast-vlan 101
huawei(config-mvlan101)#igmp program add ip 224.1.1.1 bandwidth 1024
```

13.6.2.6 (Optional) Configuring Multicast Preview

Multicast preview is an advertising method provided by carriers for ISPs. The purpose is to allow users to have an overview of a program in a controlled way. In other words, the duration, interval, and count of the user previews are controlled.

Prerequisites

The program matching mode of the multicast VLAN must be the static configuration mode.

Context

The difference between program preview and normal program watching is that, after the user goes online, the duration of the preview is restricted. When the duration expires, the user goes offline. The user can request the program again only after the preview interval expires. The count by which the user can request the program within a day (the start time can be configured) is restricted by the preview count of the user.

Multicast preview parameters are managed through the preview profile. One program can be bound to only one preview profile, but one preview profile can be referenced by multiple programs.

Table 13-34 lists the default settings of the multicast preview parameters.

 Copyright © Huawei Technologies Co., Ltd.

Table 13-34 Default settings of the multicast preview parameters

Parameter	Default Value
Global multicast preview function	enable
Preview profile	Preview profile with index 0
Preview profile parameters	Maximum preview duration: 120s Maximum preview count: 8 Minimum interval between two previews: 120s
Time for resetting the preview record	4:00:00 am
Valid duration of multicast preview	30s

Procedure

Step 1 In the global config mode, run the **btv** command to enter the BTV mode.

Step 2 Enable the global multicast preview function.

By default, the global multicast preview function is enabled. You can run the **igmp preview{ enable | disable }** command to change the setting.

Step 3 Configure the preview profile.

Run the **igmp preview-profile add** command to configure the preview profile, and set the parameters: maximum preview duration, maximum preview count, and minimum interval between two previews. The system has a default preview profile with index 0.

Step 4 Bind the program to the preview profile.

In the multicast VLAN mode, run the **igmp program add ip ip-addr preview-profile index** command to bind the program to be previewed to the preview profile so that the program has the preview attributes as defined in the preview profile. By default, the program is bound to the preview profile with index 0.

Step 5 Change the time for resetting the preview record.

Run the **igmp preview auto-reset-time** command to change the time for resetting the preview record. The preview record of the user remains valid within one day. On the second day, the preview record is reset. By default, the system resets the preview record at 4:00:00 a.m.

Step 6 Modify the valid duration of multicast preview.

Run the **igmp proxy recognition-time** or **igmp preview recognition-time** command to modify the valid duration of multicast preview. If the actual preview duration of the user is shorter than the valid duration, the preview is not regarded as a valid one and is not added to the preview count. By default, the valid duration of multicast preview is 30s.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

If you use **igmp proxy recognition-time** and **igmp preview recognition-time** commands to set the valid duration of multicast preview concurrently, the one set by the **igmp preview recognition-time** command takes effect.

Step 7 Run the **display igmp config global** command to check whether the values of the multicast preview parameters are correct.

----End

Example

To enable preview of multicast programs by using the system default preview profile, do as follows:

```
huawei(config)#btv
huawei(config-btv)#igmp preview enable
```

To enable preview of multicast programs, create preview profile 1, set the maximum preview time to 150s, the maximum preview count to 10, and apply this preview profile when adding program 224.1.1.1, do as follows:

```
huawei(config)#btv
huawei(config-btv)#igmp preview enable
huawei(config-btv)#igmp preview-profile add index 1 duration 150 times 10
huawei(config-btv)#quit
huawei(config)#multicast-vlan 101
huawei(config-mvlan101)#igmp program add ip 224.1.1.1 preview-profile 1
```

13.6.2.7 (Optional) Configuring Program Prejoin

In program prejoin, the MA5800 receives in advance the multicast stream of a program from the upper-layer multicast router to the upstream port before a user sends a request to join a program, shortening the waiting time of the user for requesting the program.

Prerequisites

The program matching mode of the multicast VLAN must be the static configuration mode.

Context

Multicast program prejoin is the same as program request. The MA5800 plays the role of a user and sends the report packet for receiving in advance the multicast stream from the upper-layer multicast router to the upstream port.

After the prejoin function is enabled, if the upper-layer multicast router does not support static multicast entry forwarding, the unsolicited report function needs to be enabled so that the user can request the program quickly. Generally, the upper-layer multicast router processes the user request by responding to the group-specific query and the general query.

NOTE

The configuration steps for IPv4 multicast and IPv6 multicast are similar but detailed commands are different. This topic describes the configuration steps for IPv4 multicast, and provides the configuration example for IPv6 multicast.

 Copyright © Huawei Technologies Co., Ltd.

Table 13-35 lists the default settings of the prejoin parameters.

Table 13-35 Default settings of the prejoin parameters

Parameter	Default Value
Prejoin function	disable
Unsolicited report of IGMP packets	disable

Procedure

Step 1 Enable the prejoin function.

Run the **igmp program add ip ip-addr prejoin enable** command to enable the prejoin function of a program. By default, the prejoin function is disabled.

Step 2 After the prejoin function is enabled, if the upper-layer multicast router does not support static multicast entry forwarding, the unsolicited report function needs to be enabled for IGMP packets.

- Run the **igmp program add ip ip-addr unsolicited enable** command to enable the unsolicited report function for IGMP packets. By default, the unsolicited report function is disabled.
- Run the **igmp unsolicited-report interval** command to modify the interval for unsolicitedly reporting IGMP packets. By default, the interval is 10s.

Step 3 Check whether the prejoin function is configured correctly.

- Run the **display igmp program** command to query the status of the prejoin function and the unsolicited report function.
- Run the **display igmp config vlan** command to query the interval for unsolicitedly reporting IGMP packets.

----End

Example

(IPv4)To enable the prejoin function when adding program 224.1.1.1 on multicast VLAN 101 for reduce the program waiting time of the users, do as follows:

```
huawei(config)#multicast-vlan 101
huawei(config-mvlan101)#igmp program add ip 224.1.1.1 prejoin enable
```

(IPv6)To enable the prejoin function when adding program ffff::1 on multicast VLAN 101 for reduce the program waiting time of the users, do as follows:

```
huawei(config)#multicast-vlan 101
huawei(config-mvlan101)#igmp program add ipv6 ffff::1 prejoin enable
```

13.6.2.8 (Optional) Configuring the Multicast Logging Function

Multicast log serves as a criterion for carriers to evaluate the viewership of multicast programs.

 Copyright © Huawei Technologies Co., Ltd.

Prerequisites

If the syslog is used for reporting multicast logs, the syslog server must be properly configured.

If the syslog server is not configured, you can run the **igmp syslog disable** command to disable the multicast syslog reporting function to save system resources.

Context

Multicast logs have three control levels: multicast VLAN level, multicast user level, and multicast program level. The system generates logs only when the logging functions at the three levels are enabled.

When the user stays online for longer than the valid time for generating logs, the system generates logs in any of the following conditions:

- The user goes offline naturally, by force, or abnormally.
- The user is blocked or deleted.
- The program is deleted.
- The program priority is changed.
- The upstream port to which the program is bound changes.
- The VLAN of the upstream port to which the program is bound changes.
- The right mode is switched.
- The user preview times out.
- The IGMP mode is switched.
- The bandwidth CAC is not passed.

When the user goes online, the system records only the online date and time. The system generates a complete log only when the user goes offline.

The MA5800 can report the multicast log to the log server in the syslog mode and the call detailed record (CDR) mode. By default, the MA5800 reports the log in the syslog mode.

- Syslog mode: Logs are reported to the syslog server in the form of a single log.
- CDR mode: Logs are reported to the log server in the form of a log file (.cvs). One log file contains multiple logs.

Table 13-36 lists the default settings of the multicast logging parameters.

Table 13-36 Default settings of the multicast logging parameters

Parameter	Default Value
Report mode of the multicast log	Syslog mode
Logging function at the multicast VLAN level	enable

 Copyright © Huawei Technologies Co., Ltd.

Parameter	Default Value
Logging function at the multicast user level	enable
Logging function at the multicast program level	enable
Action report function of the multicast user	disable
Interval for automatically logging	2 hours
Minimum online duration for generating a valid log	30s
Parameters of the log report in the CDR mode	Report interval: 600s Maximum number of logs that can be reported each time: 200

Procedure

- Configure the parameters of the logging function of the multicast host.
  a. Enable the multicast logging functions.

     Multicast logs have three control levels: multicast VLAN level, multicast user level, and multicast program level. The system generates logs only when the logging functions at the three levels are enabled. By default, the three functions are enabled.

     - In the Multicast VLAN mode, run the **igmp log { enable | disable }** command to configure the logging function at the multicast VLAN level.
     - In the BTV mode, run the **igmp user add service-port index log { enable | disable }** command to configure the logging function at the multicast user level.

       In the BTV mode, run the **igmp log record { user | mac }** command to configure the log record object. After the configuration, the device can record ordering action of users or multicast terminals identified by MAC addresses.
     - In the Multicast VLAN mode, run the **igmp program add ip ip-addr log { enable | disable }** command to configure the logging function at the multicast program level.

  b. Modify the interval for automatically logging.

     In the BTV mode, run the **igmp proxy log-interval** command to modify the interval for automatically logging. When the user stays online for a long time, the system generates logs at the preset interval. This is to prevent the problem that a log is not generated when the user leaves the multicast group without sending a leave packet, which can affect the accounting. By default, the interval is two hours.

 Copyright © Huawei Technologies Co., Ltd.

c. Modify the minimum online duration for generating a valid log.

In the BTV mode, run the **igmp proxy recognition-time** or **igmp log recognition-time** command to modify the minimum online duration for generating a valid log. If the user is in a multicast group (such as to preview a program) for shorter than the preset duration, the user operation is not regarded as a valid one and a log is not generated. A log is generated only when a user stays online for longer than the specified duration. By default, the minimum online duration is 30s.

NOTE

If you use **igmp proxy recognition-time** and **igmp log recognition-time** commands to set the minimum online duration for generating a valid log concurrently, the one set by the **igmp log recognition-time** command takes effect.

- (Optional) Configure the action report function of the multicast user.

  By default, the system uses the syslog mode to report multicast logs. You can run the **igmp user-action-report** command to configure the action report function of the multicast user. By default, the action report function of the multicast user is disabled.

  – **enable**: Enables the action report function of the multicast user. Logs are reported to the syslog server when a multicast user goes online and offline.
  – **disable**: Disables the action report function of the multicast user. Logs are reported to the syslog server only when a multicast user goes offline.

- Configure the function of CDR-mode log report.
  a. Configure the multicast log server and the data transmission mode for the CDR-mode log report.

     Run the **file-server auto-backup cdr** command to configure the active and standby multicast log servers.
  b. Enable the function of CDR-mode log report.

     In the BTV mode, run the **igmp cdr { enable | disable }** command to configure the function of CDR mode log report. After the function is enabled, the MA5800 reports the local multicast logs to the multicast log server in the form of a file. After the function is disabled, the MA5800 reports each single log to the syslog server in the default syslog mode.
  c. Configure the parameters of the log report in the CDR mode.

     ▪ In the BTV mode, run the **igmp cdr-interval** command to set the report interval. By default, the interval is 600s.

     ▪ In the BTV mode, run the **igmp cdr-number** command to set the maximum number of logs that can be reported each time. When the number of the multicast logs in the CDR file reaches the preset value, the MA5800 reports the logs. By default, the maximum number is 200.
  d. Check whether the configuration is correct.

     ▪ Run the **display file-server** command to query the configuration of the CDR multicast log server.

 Copyright © Huawei Technologies Co., Ltd.

- Run the **display igmp config global** command to query the status and other parameters of the function of CDR-mode log report.

----End

Example

To configure the multicast log to be reported to log server 10.10.10.1 in the CDR mode, and use the SFTP transmission mode, do as follows:

```
huawei(config)#file-server auto-backup cdr primary 10.10.10.1 sftp path temp user
huawei(config)#btv
huawei(config-btv)#igmp cdr enable
```

13.6.2.9 (Optional) Configuring the Maximum Number of Programs That Can Be Watched by the Multicast User

This topic describes how to configure the maximum number of programs that can be ordered by the multicast user at the same time. You can configure the maximum number of all programs that can be watched by the multicast user at the same time, or configure the maximum number of the different-level programs that can be watched by the multicast user.

Prerequisites

When you configure the maximum number of programs based on the program level, the program level must be configured at the same time and the programs must be configured statically.

Context

Table 13-37 lists the default settings of the max-program parameters.

Table 13-37 Default settings of the multicast max-program parameters

Parameter	Default Value
Maximum number of programs that can be watched by the multicast user	8
Grade of the multicast program	no-grade
Maximum number of programs of different priorities that can be watched by the multicast user	no-limit

 Copyright © Huawei Technologies Co., Ltd.

Procedure

Step 1 In the global config mode, run the **btv** command to enter the BTV mode.

Step 2 Configure the max-program of the multicast user.

Run the **igmp user add service-port index max-program max-program-num** command to set the maximum number of programs that can be watched by the multicast user.

Step 3 Configure the maximum number of programs of different priorities that can be watched by the multicast user.

Run the **igmp user watch-limit service-port index** command to set the maximum number of programs of different priorities that can be watched by the multicast user.

Step 4 Configure the grade of the multicast program.

In the multicast VLAN mode, run the **igmp program add ip ip-addr grade** command to configure the grade of a multicast program.

Step 5 Check whether the multicast max-program configuration is correct.

- Run the **display igmp user** command to query the maximum number of programs that can be watched and watching by the multicast user.
- Run the **display igmp program** command to query the grade of the multicast program.
- Run the **display igmp user extended-attributes service-port** command to query the maximum number of programs that can be watched and watching by the multicast user.

----End

Example

To set the user max-program to 10 when adding multicast service-port 0, set the user can watch 2 HDTV program, and configure the program grade to hdtv when adding multicast program 224.1.1.1, do as follows:

```
huawei(config)#btv
huawei(config-btv)#igmp user add service-port 0 max-program 10
huawei(config-btv)#igmp user watch-limit service-port 0 hdtv 2
huawei(config-btv)#quit
huawei(config)#multicast-vlan 101
huawei(config-multicast-vlan 101)#igmp program add ip 224.1.1.1 grade hdtv
```

13.6.2.10 (Optional) Configuring the Maximum Rate for Sending IGMP Packets

When the multicast and anti-DoS attack functions are enabled, the system reports DoS attack alarms and drops IGMP packets over the rate limit, if a user port sends such IGMP packets to the CPU. When the anti-DoS attack function is disabled, the system always sends IGMP packets to the CPU. If they are sent to the CPU at a rate higher than the limit, the system drops the IGMP packets that exceed the rate limit but does not report DoS attack alarms.

 Copyright © Huawei Technologies Co., Ltd.

Prerequisites

The multicast function is enabled.

Context

When the multicast function is enabled, the system will always send the received IGMP packets to the CPU if no control is implemented over the process. Then if a user port receives a large number of IGMP packets, the IGMP packets for other users will not be processed and directly dropped instead.

Table 13-38 lists the default maximum rate settings for sending user IGMP packets to the CPU.

Table 13-38 Default rate settings for sending user IGMP packets to the CPU

Parameter	Default Value
Anti-DoS attack function	disable
Maximum rate for sending IGMP packets to the CPU	63 pps

Procedure

Step 1 Enable the anti-DoS attack function.

Run the **security anti-dos { enable | disable }**command to enable the anti-DoS attack function, which is disabled by default.

Step 2 Specify the maximum rate for sending IGMP packets to the CPU.

Run the **security anti-dos control-packet igmp rate frameid/slotid/portid { value | ont | no-limit** command to specify the maximum rate for sending IGMP packets to the CPU, which is 63 by default.

----End

Example

Example: Specify the maximum rate for sending IGMP packets to the CPU to 63 pps for user port 0/2/1 and enable the system to report the port to the blacklist if it sends IGMP packets over the rate limit to the CPU.

```
huawei(config)#security anti-dos enable
huawei(config)#security anti-dos control-packet igmp rate 0/2/1  ont 0 gemport 2 20
```

13.6.3 Configuring the Multicast Service on a Cascading Network

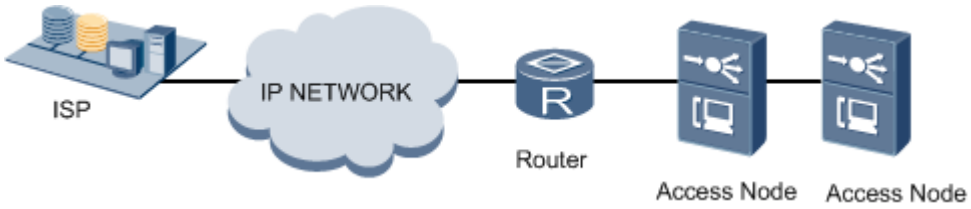
This topic describes how to configure the multicast service on the MA5800 on a cascading network.

 Copyright © Huawei Technologies Co., Ltd.

Application Context

Figure 13-61 shows the application context of the multicast service in a cascading network. When a cascaded device needs to provision the multicast service, the cascading port on the cascading device needs to be configured as a multicast cascading port. In this way, the cascading device regards the cascaded device as an IGMP user.

Figure 13-61 Application context of multicast service in a cascading network



Default Configuration

Table 13-39 lists the default settings of the multicast service of the MA5800.

Table 13-39 Default settings of the multicast service

Feature	Default Settings
Multicast protocol	off (multicast function disabled)
IGMP version, includes multicast user version and multicast VLAN version.	V3
Multicast program configuration mode	Static configuration mode
Multicast bandwidth management	enabled
Multicast preview	enabled
Multicast log function	enabled

Precaution

The multicast program list of the cascading device must cover the multicast program list of the cascaded device.

Procedure

The procedure for configuring the cascaded device is the same as described in **Configuring the Multicast Service in a Single-NE Network**.

The procedure of configuring the cascading device is as follows:

1. For details on configuring the multicast service, see **Configuring the Multicast Service in a Single-NE Network**.

 Copyright © Huawei Technologies Co., Ltd.

2. Configure the multicast cascading port.

   Run the **igmp cascade-port frameid/slotid/portid** command to configure the cascading port as the multicast cascading port. The multicast upstream port cannot be configured as a multicast cascading port.

3. When the cascaded device requires the quick leave function of the multicast user, run the **igmp cascade-port frameid/slotid/portid quickleave enable** command to enable the quick leave attribute on the multicast cascading port.

NOTICE

If the lower-layer device does not support the proxy of the IGMP leave packet, all the users requesting the program may go offline when a user requesting the same program goes offline. Therefore, when the quick leave attribute is enabled on the multicast cascading port, it is recommended that the lower-layer device use the IGMP proxy function, or switch to the IGMP snooping mode with the IGMP leave packet proxy function enabled.

13.6.4 Configuring the Multicast Service in an MSTP Network

This topic describes how to configure the multicast service in an MSTP network.

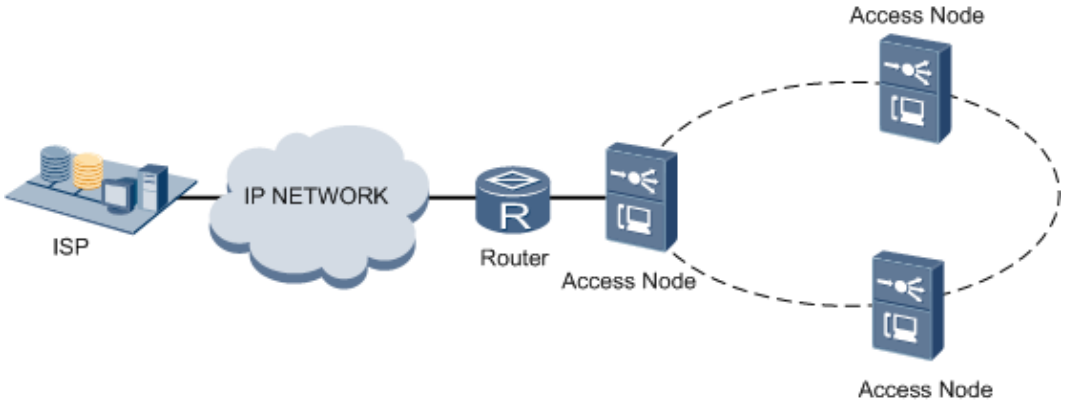
Prerequisites

Basic configurations for the MSTP network are complete. For details about configuring the MSTP network, see **Configuring the MSTP**.

Application Context

Figure 13-62 shows the application context of the multicast service in an MSTP network. When the multicast service is provisioned in an MSTP ring network, the multicast upstream port and the cascading port need to be added to the multicast VLAN. According to the running result of the MSTP protocol, the multicast request packets are sent from the root port or the default port (when the device is a root bridge), and the other ports in the VLAN serve as cascading ports.

Figure 13-62 Application context of the multicast service in an MSTP network



Copyright © Huawei Technologies Co., Ltd.

Default Configuration

Table 13-40 Default settings of the multicast service

Feature	Default Settings
Multicast protocol	off (multicast function disabled)
IGMP version, includes multicast user version and multicast VLAN version.	V3
Multicast program configuration mode	Static configuration mode
Multicast bandwidth management	enabled
Multicast preview	enabled
Multicast log function	enabled

Procedure

For details on configuring the MSTP ring network, see Configuring the MSTP.

1. For details on configuring the multicast service, see **Configuring the Multicast Service in a Single-NE Network**.
2. Configure the MSTP multicast upstream port.

   When multicast service is provisioned in an MSTP network, the multicast upstream port needs to be set to the MSTP mode, and the default upstream port of the multicast VLAN can be specified. After the configuration is completed, multicast packets are forwarded by the root port or default port of the multicast VLAN.

   - Run the **igmp uplink-port-mode mstp** command to set the upstream port to the MSTP mode.
   - Run the **igmp default uplink-port** command to specify the default upstream port of the multicast VLAN. When the upstream port is set to the MSTP mode and an MSTP root port is not available in the multicast VLAN, the multicast VLAN by default adopts the upstream port as the multicast upstream port.
3. Configure the multicast cascading port.

   Run the **igmp cascade-port** command to configure the cascading port as the multicast cascading port.
4. Configure multicast quick convergence in the case of an MSTP network topology change.

   Multicast quick convergence means that the device can quickly join the multicast group through a new upstream port when the MSTP network topology changes. The device can unsolicitedly send the new upstream port the IGMP join packet for an online program so that the device joins all the multicast groups. Or, the device can send the IGMP global leave packet to the upstream port. Then, the upper-layer querier sends a query packet for generating a new multicast forwarding tree.

 Copyright © Huawei Technologies Co., Ltd.

Run the **igmp send global-leave** command to enable the function of sending the IGMP global leave packet. When this function is enabled, the device sends the IGMP global leave packet to the upper-layer multicast router. When this function is disabled, the device sends the IGMP join packet to the upper-layer multicast router. By default, the function of sending the IGMP global leave packet is enabled.

13.6.5 Configuring the PIM Service

This topic describes how to configure the Protocol Independent Multicast (PIM) service.

Prerequisites

- The access network is enabled with routing and can obtain the rendezvous point (RP) or routing information about an ISP device.
- The upper-layer device interconnected with the access device is enabled with PIM.

Context

A multicast network requires multicast protocols to replicate and forward multicast data. PIM is a widely used intra-domain multicast protocol that builds MDTs to transmit multicast data between routers in the same domain.

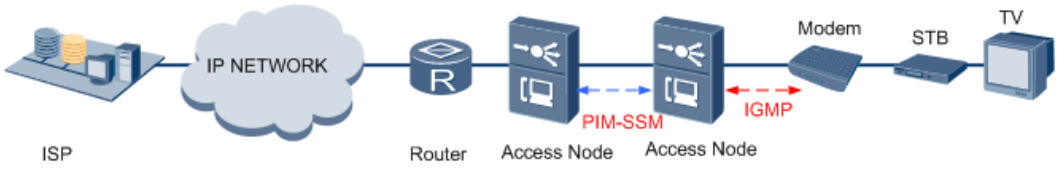
PIM can create multicast routing entries on demand, forward packets based on multicast routing entries, and dynamically respond to network topology changes.

Concurrently, the access node supports PIM-SM (used in an IPv4 network) and PIM-SSM (used in IPv4 and IPv6 networks).

Application Context

Figure 13-63 shows the PIM networking. The access node exchanges with the upper-layer device through PIM and exchanges with the user-side device through IGMP.

Figure 13-63 PIM networking



Procedure

Step 1 Create a multicast VLAN and VLAN of the upstream PIM interface, and add the upstream port to the VLAN of the PIM interface.

1. Run the **vlan** command to create a multicast VLAN and VLAN of the upstream PIM interface.

 Copyright © Huawei Technologies Co., Ltd.

2. Run the **port vlan vlan-list frameid/slotid portlist** to add the upstream port to the VLAN of the PIM interface.

Step 2 Enable the multicast function.

Run the **multicast routing-enable** command to enable multicast or run the **multicast ipv6 routing-enable** command to enable IPv6 multicast.

Step 3 Set upstream multicast mode to PIM.

- If PIM-SM is used in the upstream direction, run the **multicast upstream-mode pim-sm** command to set the upstream multicast exchange mode to PIM-SM.
- If PIM-SSM is used in the upstream direction, run the **multicast upstream-mode pim-ssm** command to set the upstream multicast exchange mode to PIM-SSM.

Step 4 Create the VLAN interface of the upstream PIM interface, configure the IP address of the Layer 3 interface, and enable PIM on this Layer 3 interface.

1. Run the **interface vlanif** command to create the VLAN interface of the upstream PIM interface.
2. Run the **ip address** or **ipv6 address** command to configure the IP addresses of the Layer 3 interface.
3. Run the **pim sm** or **pim ipv6 sm** command to enable PIM on the Layer 3 interface.

Step 5 Create the multicast VLAN interface, configure the IP address of the Layer 3 interface, and enable multicast on this Layer 3 interface.

1. Run the **interface vlanif** command to create a multicast VLAN interface.
2. Run the **ip address** or **ipv6 address** command to configure the IP addresses of the Layer 3 interface.

Step 6 (Optional) If PIM-SM is enabled in the upstream direction, static RPs need to be configured.

Run the **static-rp** command to configure static RPs.

Step 7 Configure multicast programs.

Run the **igmp program add** command to configure a multicast program.

Step 8 Configure multicast users.

Run the **igmp user add** command to configure a multicast user.

----End

Example

To enable PIM-SSM in the upstream direction and IGMP in the downstream direction, do as follows:

- Upstream port: 0/9/0
- VLAN of the upstream PIM-SSM interface: 200; IP address of the Layer 3 interface: 10.1.1.1; mask: 24
- Multicast VLAN: 100; IP address of the Layer 3 interface: 10.2.2.2; mask: 24

 Copyright © Huawei Technologies Co., Ltd.

- Service VLAN 100 for the multicast service; all users in the VLAN are added as multicast users in batches; Set all multicast users in the service VLAN as members of the multicast VLAN.
- Multicast program: static; program name: movie; program IP address: 232.1.1.1; source IP address: 10.1.1.2; program bandwidth: 5000 kbit/s

```
huawei(config)#vlan 100
huawei(config)#vlan 200
huawei(config)#port vlan 200 0/9 0
huawei(config)#multicast routing-enable
huawei(config)#multicast upstream-mode pim-ssm
huawei(config)#interface vlanif 200
huawei(config-if-vlanif200)#ip address 10.1.1.1 24
huawei(config-if-vlanif200)#pim sm
huawei(config-if-vlanif200)#quit
huawei(config)#interface vlanif 100
huawei(config-if-vlanif100)#ip address 10.2.2.2 24
huawei(config-if-vlanif100)#quit
huawei(config)#btv
huawei(config-btv)#igmp user add smart-vlan 100
huawei(config-btv)#quit
huawei(config)#multicast-vlan 100
huawei(config-mvlan100)#igmp multicast-vlan member smart-vlan 100
huawei(config-mvlan100)#igmp program add name movie ip 232.1.1.1 sourceip 10.1.1.2 bandwidth 5000
huawei(config-mvlan100)#quit
```

13.6.6 Configuring the Layer 2 Reverse Multicast Service

This topic describes how to configure the Layer 2 reverse multicast service.

Prerequisites

- An OLT has been installed.
- The IP route between the OLT and the network-side devices is reachable. The OLT and ONUs can be managed by eSight/Cloud Campus.
- The link between the OLT and ONUs is normal, and the ONUs can register with the OLT and go online.

Context

Generally, the upper-layer devices of an OLT, such as video servers, are considered as multicast sources. A video server issues the service stream of multicast programs to users from top down. This is forward multicast. However, in some scenarios, for example, in a subway station with a surveillance center and multiple sub-surveillance centers, reverse multicast is required when the stream of some cameras needs to be displayed on several large screens.

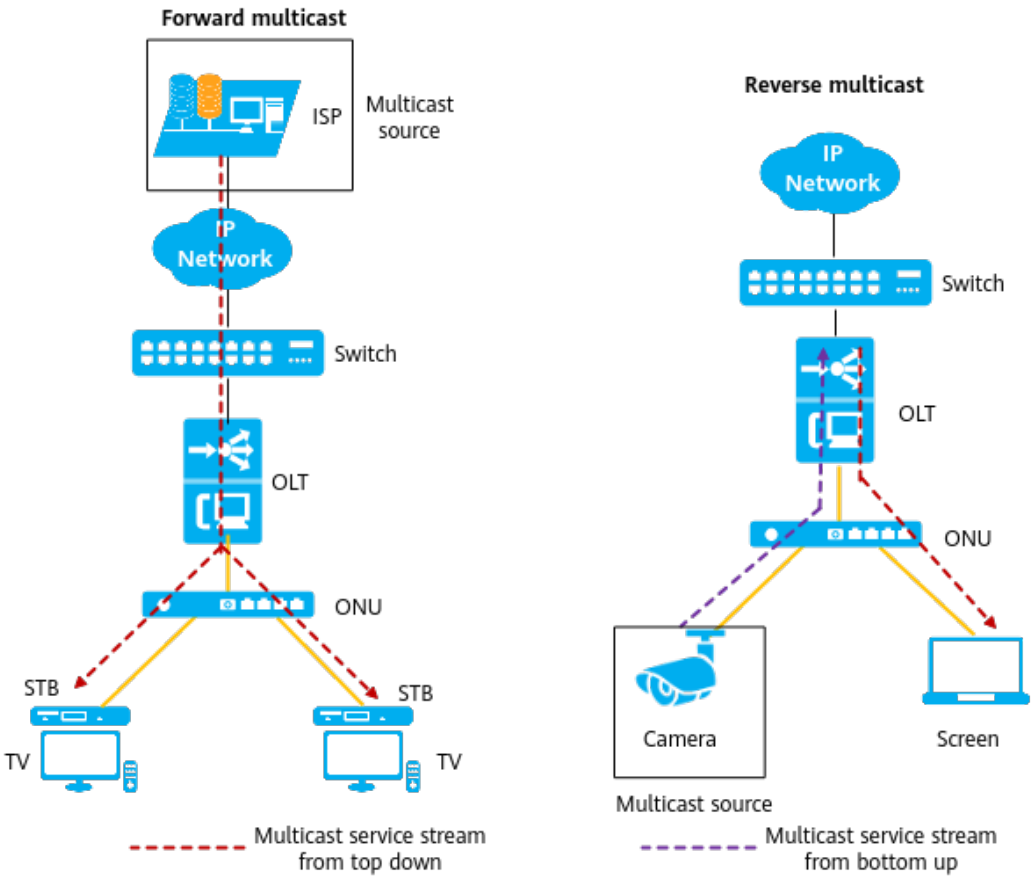
In reverse multicast, the multicast source comes from a user under the OLT. The service stream from the multicast source is replicated on the OLT and then sent to other multicast users in the same multicast VLAN. **Figure 13-64** shows forward multicast and reverse multicast services.

NOTE

In the reverse multicast scenario, the multicast source must be directly connected to an ONU. Layer 2 or Layer 3 multicast networking is not supported under the ONU.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-64 Reverse multicast diagram

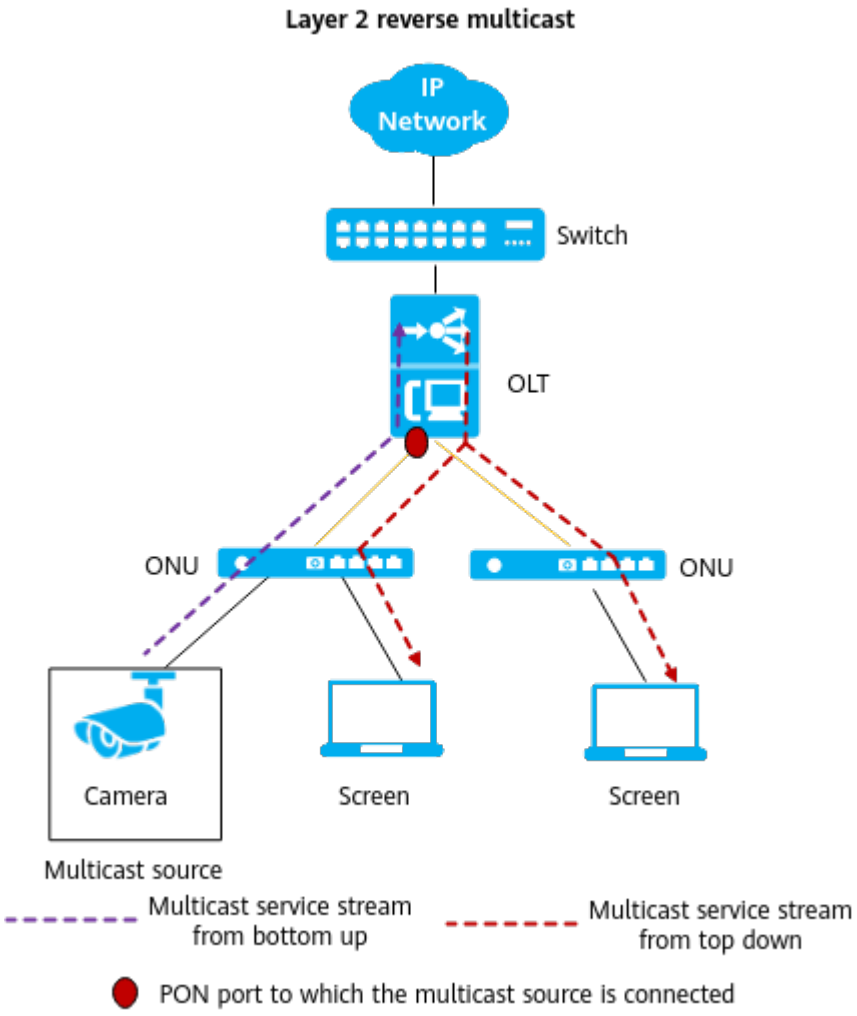


Application Environment

Figure 13-65 shows the application of Layer 2 reverse multicast. Users receiving the multicast service stream can be connected to the same ONU or different ONUs.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-65 Layer 2 reverse multicast application diagram



Procedure

Step 1 Enable the reverse multicast function.

Run the **multicast-source port** command to enable reverse multicast.

Step 2 Enable the ONU reverse multicast stream forwarding function.

Run the **port multicast-source eth *ont-portlist* enable** command to enable an ONU to forward reverse multicast stream in an ONU service profile.

Step 3 Create a multicast VLAN, and configure multicast programs and multicast users.

- Run the **multicast-vlan** command to create a multicast VLAN.
- Run the **igmp program add** command to configure multicast programs.
- Run the **igmp user add** command to configure multicast users.

Step 4 Configure reverse multicast service ports.

1. Configure C-VLAN = S-VLAN = M-VLAN for multicast users on the PON port directly connected to the multicast source. The C-VLAN and S-VLAN can be

 Copyright © Huawei Technologies Co., Ltd.

set using the **service-port** *service-port* **vlan** *SVLAN* **gpon** *frameid/slotid/portid* **ont** *ontid* **gemport** *gemindex* **multi-service user-vlan** *CVLAN* command, and the M-VLAN can be set using the **multicast-vlan** command.

NOTE

The multicast source ingress of the same multicast program must be unique, either a forward multicast source ingress or a reverse multicast source ingress.

----End

Example

Configure Layer 2 reverse multicast for a single device with the following configurations:

- The multicast source is directly connected to port eth 1 on ONU 0 which is connected to PON port 0/2/0. Set the service profile ID to 100, service port ID to 0, and S-VLAN and C-VLAN IDs to 100.
- Set the multicast VLAN ID to 100, multicast program mode to static, program name to movie, program IP address to 231.1.1.1, source IP address to 10.1.1.2, and program bandwidth to 5000 Kbit/s.
- Plan service VLAN 100 for the multicast service, and add all users in the VLAN as multicast users in batches.

```
huawei(config)#btv
huawei(config-btv)#multicast-source port 0/2 0
huawei(config-btv)#quit
huawei(config)#ont-srvprofile gpon profile-id 100
huawei(config-gpon-srvprofile-100)#port multicast-source eth 1 enable
huawei(config-gpon-srvprofile-100)#commit
huawei(config-gpon-srvprofile-100)#quit
huawei(config)#vlan 100
huawei(config)#service-port 0 vlan 100 gpon 0/2/0 ont 0 gemport 0 multi-service user-vlan 100
huawei(config)#btv
huawei(config-btv)#igmp user add smart-vlan 100
huawei(config-btv)#quit
huawei(config)#multicast-vlan 100
huawei(config-mvlan100)#igmp program add name movie ip 231.1.1.1 sourceip 10.1.1.2 bandwidth 5000
huawei(config-mvlan100)#igmp multicast-vlan member smart-vlan 100
```

13.6.7 Configuring the Layer 3 Reverse Multicast Service

This topic describes how to configure the Layer 3 reverse multicast service.

Prerequisites

- An OLT has been installed.
- The IP route between the OLT and the network-side devices is reachable. The OLT and ONUs can be managed by eSight/Cloud Campus.
- The link between the OLT and ONUs is normal, and the ONUs can register with the OLT and go online.

Context

Generally, the upper-layer devices of an OLT, such as video servers, are considered as multicast sources. A video server issues the service stream of multicast

 Copyright © Huawei Technologies Co., Ltd.

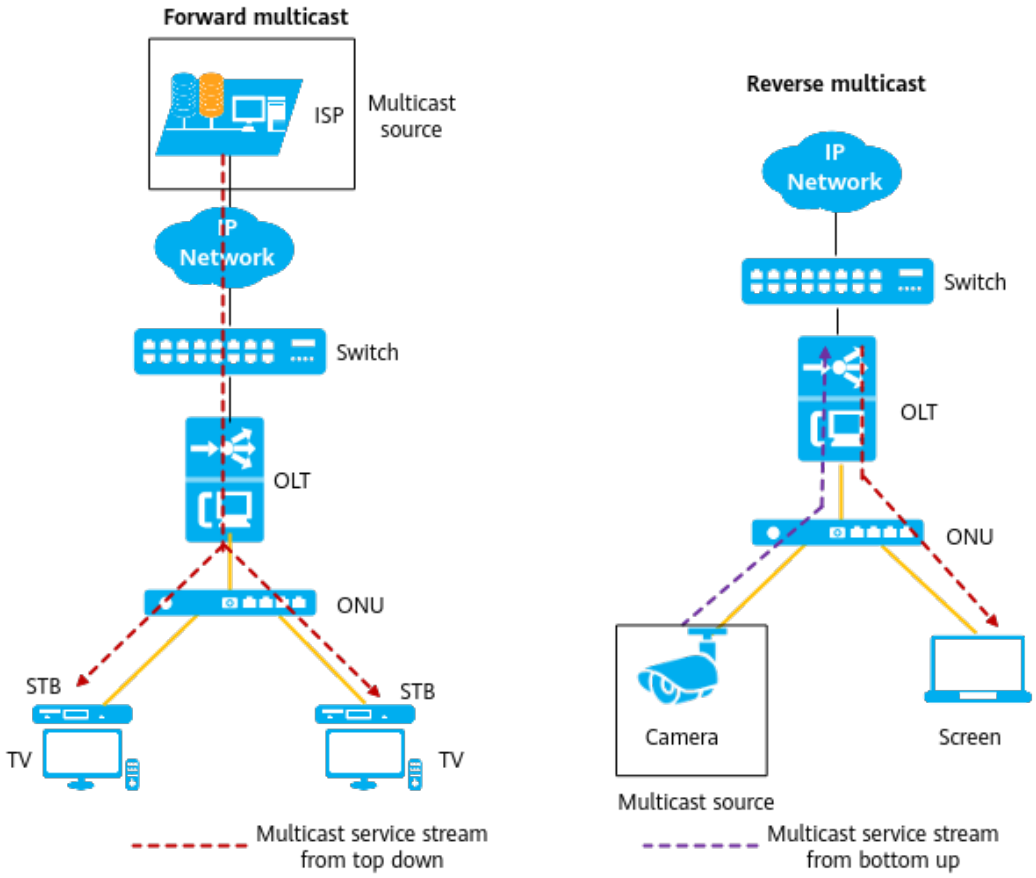
programs to users from top down. This is forward multicast. However, in some scenarios, for example, in a subway station with a surveillance center and multiple sub-surveillance centers, reverse multicast is required when the stream of some cameras needs to be displayed on several large screens.

In reverse multicast, the multicast source comes from a user under the OLT. The service stream from the multicast source is replicated on the OLT and then sent to other multicast users in the same multicast VLAN. **Figure 13-66** shows forward multicast and reverse multicast services.

NOTE

- In the reverse multicast scenario, the multicast source must be directly connected to an ONU. Layer 2 or Layer 3 multicast networking is not supported under the ONU.
- You are not allowed to set the group filter mode in a PIM-enabled multicast VLAN.

Figure 13-66 Reverse multicast diagram

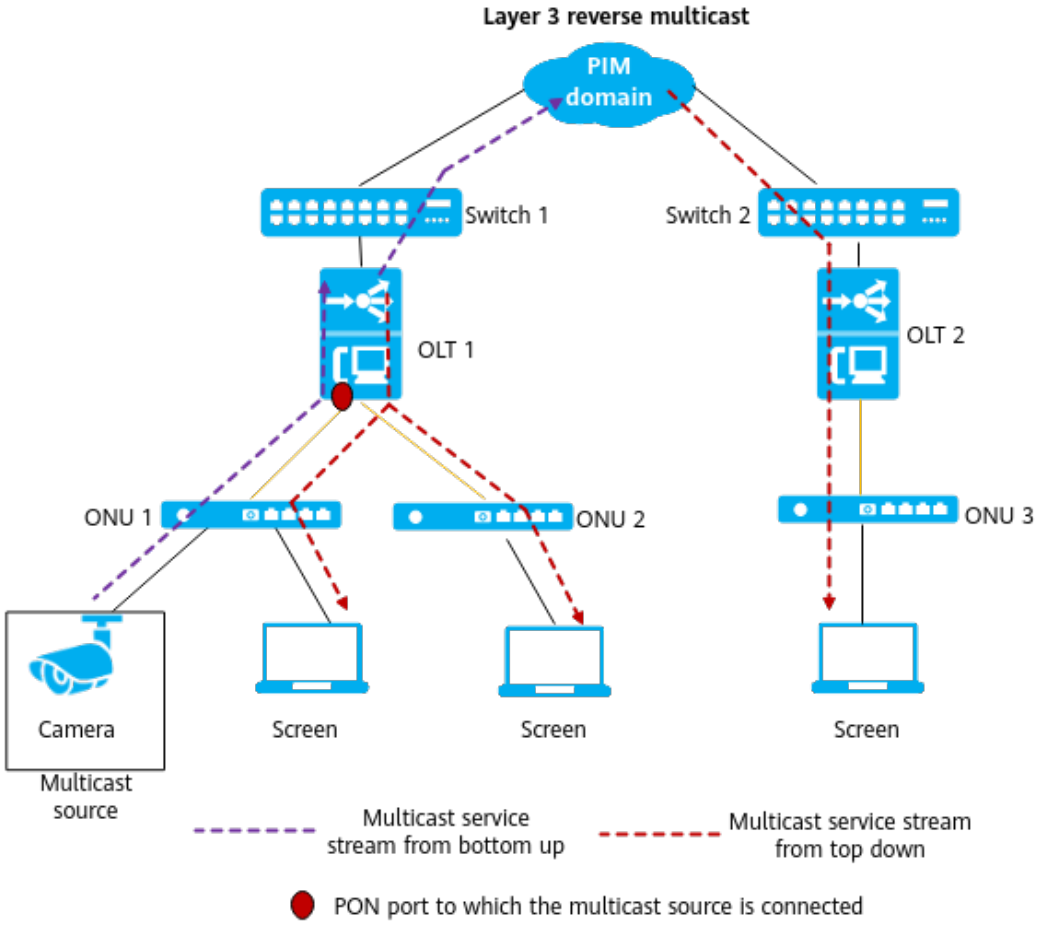


Application Environment

Figure 13-65 shows the application of Layer 3 reverse multicast. Users receiving the multicast service stream are connected to different OLTs.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-67 Layer 3 reverse multicast application diagram



Procedure

Step 1 Enable the reverse multicast function.

Run the **multicast-source port** command to enable reverse multicast.

Step 2 Enable the ONU reverse multicast stream forwarding function.

Run the **port multicast-source eth *ont-portlist* enable** command to enable an ONU to forward reverse multicast stream in an ONU service profile.

Step 3 Create a multicast VLAN and an upstream PIM interface VLAN, and add the PIM router port to the PIM interface VLAN.

1. Run the **vlan** command to create a multicast VLAN and upstream PIM interface VLAN.
2. Run the **port vlan vlan-list frameid/slotid portlist** command to add the PIM router port to the PIM interface VLAN.

Step 4 Enable the multicast function.

Run the **multicast routing-enable** command to enable the multicast function.

 Copyright © Huawei Technologies Co., Ltd.

Step 5 Set the multicast upstream mode to PIM.

- If PIM-SM is used in the upstream direction, run the **multicast upstream-mode pim-sm** command to set the upstream multicast exchange mode to PIM-SM.
- If PIM-SSM is used in the upstream direction, run the **multicast upstream-mode pim-ssm** command to set the upstream multicast exchange mode to PIM-SSM.

Step 6 Create the VLAN Layer 3 interface of the upstream PIM interface, configure the IP address of the Layer 3 interface, and enable PIM on this Layer 3 interface.

1. Run the **interface vlanif** command to create the VLAN Layer 3 interface for the upstream PIM interface.
2. Run the **ip address** command to configure the IP address of the Layer 3 interface.
3. Run the **pim sm** command to enable PIM on the Layer 3 interface.

Step 7 Create a multicast VLAN Layer 3 interface, configure the IP address of the Layer 3 interface, and enable the multicast function on the Layer 3 interface.

1. Run the **interface vlanif** command to create a multicast VLAN Layer 3 interface.
2. Run the **ip address** command to configure the IP address of the Layer 3 interface.

Step 8 (Optional) If PIM-SM is enabled in the upstream direction, static rendezvous points (RPs) need to be configured.

Run the **static-rp** command to configure static RPs.

Step 9 Configure multicast programs and multicast users.

- Run the **igmp program add** command to configure multicast programs.
- Run the **igmp user add** command to configure multicast users.

Step 10 Configure reverse multicast service ports.

- Configure C-VLAN = S-VLAN = M-VLAN for multicast users on the PON port directly connected to the multicast source. The C-VLAN and S-VLAN can be set using the **service-port** *service-port* **vlan** *SVLAN* **gpon** *frameid/slotid/portid* **ont** *ontid* **gemport** *gemindex* **multi-service user-vlan** *CVLAN* command, and the M-VLAN can be set using the **multicast-vlan** command.

NOTE

The multicast source ingress of the same multicast program must be unique, either a forward multicast source ingress or a reverse multicast source ingress.

----End

Example

When PIM-SSM is used in the upstream direction and IGMP is used in the downstream direction for the multicast service, configure Layer 3 reverse multicast with the following configurations:

 Copyright © Huawei Technologies Co., Ltd.

- The multicast source is directly connected to port eth 1 on ONU 0 which is connected to PON port 0/2/0. Set the service profile ID to 100 and service port ID to 0.
- The PIM router port is 0/9/0.
- On the PIM router side, the VLAN of the PIM-SSM interface is 200, the IP address of the Layer 3 interface is 10.1.1.1, and the mask is 24. The PIM router side sets up the Layer 3 multicast relationship with the OLT by sending PIM packets carrying multicast sources.
- The multicast VLAN ID is 100, the IP address of the Layer 3 interface is 10.2.2.2, and the mask is 24.
- Set the multicast program mode to static, program name to movie, program IP address to 232.1.1.1, source IP address to 10.1.1.2, and program bandwidth to 5000 Kbit/s.
- Plan service VLAN 100 for the multicast service, and add all users in the VLAN as multicast users in batches.

```
huawei(config)#btv
huawei(config-btv)#multicast-source port 0/2 0
huawei(config-btv)#quit
huawei(config)#ont-srvprofile gpon profile-id 100
huawei(config-gpon-srvprofile-100)#port multicast-source eth 1 enable
huawei(config-gpon-srvprofile-100)#commit
huawei(config-gpon-srvprofile-100)#quit
huawei(config)#vlan 100
huawei(config)#vlan 200
huawei(config)#port vlan 200 0/9 0
huawei(config)#service-port 0 vlan 100 gpon 0/2/0 ont 0 gemport 0 multi-service user-vlan 100
huawei(config)#multicast routing-enable
huawei(config)#multicast upstream-mode pim-ssm
huawei(config)#interface vlanif 200
huawei(config-if-vlanif200)#ip address 10.1.1.1 24
huawei(config-if-vlanif200)#pim sm
huawei(config-if-vlanif200)#quit
huawei(config)#interface vlanif 100
huawei(config-if-vlanif100)#ip address 10.2.2.2 24
huawei(config-if-vlanif100)#quit
huawei(config)#multicast-vlan 100
huawei(config-mvlan100)#igmp program add name movie ip 232.1.1.1 sourceip 10.1.1.2 bandwidth
5000
huawei(config-mvlan100)#quit
huawei(config)#btv
huawei(config-btv)#igmp user add smart-vlan 100
```

13.7 Multicast Maintenance and Diagnosis

13.7.1 Multicast Emulation

A multicast emulation test remotely emulates an end user going online, and engineers query the real-time traffic of the multicast program to determine whether the multicast function is running properly.

13.7.1.1 Introduction

In multicast emulation, an access device remotely emulates an end user going online. Engineers query the real-time traffic of the multicast program to determine whether the multicast function is running properly.

 Copyright © Huawei Technologies Co., Ltd.

Multicast emulation is used in acceptance tests or fault location. The following table lists the comparison between the multicast emulation test and traditional tests.

Table 13-41 Comparison between the multicast emulation test and traditional tests

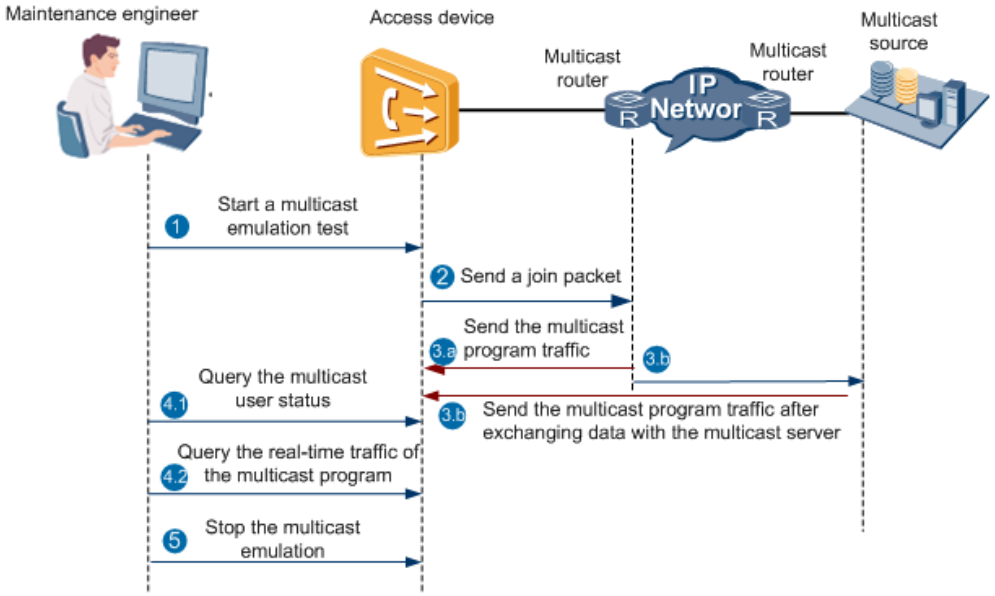
Scenario	Task	Traditional Test	Multicast Emulation
Acceptance test	After an access device is installed and configured with data, a test engineer needs to check whether the multicast service has been provisioned to the access device successfully.	The test engineer visits the site where the access device is installed and uses an external tester or a portable computer to perform a multicast test on each port of the device.	The test engineer remotely logs in to the access device to perform a multicast emulation test and determines the service status based on the test results.
Fault location	The multicast service is abnormal, and a maintenance engineer needs to quickly locate the network segment of the fault to facilitate subsequent troubleshooting.	The maintenance engineer visits all sites where the access devices are installed and performs tests.	The maintenance engineer remotely logs in to an access device and performs a multicast emulation test to preliminarily determine the network segment of the fault. Based on the test results, the engineer diagnoses the fault cause and rectifies the fault. **NOTE** A multicast emulation test cannot check the status of the line between an end user and an access device.

13.7.1.2 Principles

Figure 13-68 shows the principles of multicast emulation.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-68 Principles of multicast emulation



1. A maintenance engineer remotely logs in to an access device and starts a multicast emulation test on a user port.

   The engineer sets parameters, such as the user information, program IP address, and multicast VLAN ID.
2. The access network device constructs a join packet and sends the packet to the multicast router for joining a multicast group.
3. The multicast router checks whether the multicast program traffic exists.
   - If the multicast program traffic exists, the multicast router sends the multicast program traffic to the access device.
   - If the multicast program traffic does not exist, the multicast source sends the multicast program traffic to the access device after exchanging data with the multicast router.
4. The maintenance engineer queries the status of the emulation user and the real-time traffic of the multicast program.
   - Checks whether the multicast emulation user is online to determine whether the user successfully orders the program.
   - Checks the real-time traffic of the multicast program to determine whether the communication between the access device and the multicast source is normal.
5. After the multicast emulation is complete, the maintenance engineer stops the emulation manually using the CLI to release resources.

 Copyright © Huawei Technologies Co., Ltd.

13.7.1.3 Usage Scenario

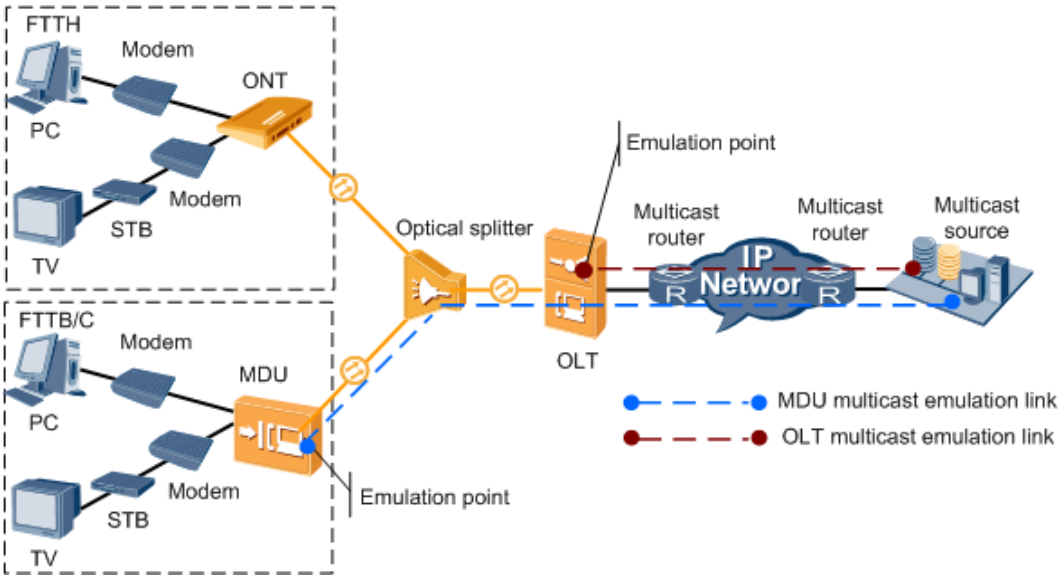
Context

On fiber to the x (FTTx) networks, access devices are located closer to user terminals and widely distributed. In a multicast emulation test, an access node emulates the multicast user to implement remote acceptance for services and locate faults, which reduces the O&M costs.

Scenario

Figure 13-69 shows the multicast emulation on a typical FTTx network.

Figure 13-69 Multicast emulation on a typical FTTx network



NOTE

In Figure 13-69, the MDU and ONT can provide the video on demand (VoD) service for a user through the PC. The MDU and ONT can also provide the BTV service using a set top box (STB).

In IPTV service acceptance or fault location:

- For MDU multicast users, multicast emulation can be performed on the MDU in service acceptance.
- For ONT multicast users, multicast emulation can be performed on the OLT.

  NOTE

  ONTs do not support multicast emulation. To emulate an ONT multicast user, perform an emulation test on the PON board of the OLT.

 Copyright © Huawei Technologies Co., Ltd.

Fault Location

After the multicast emulation, you can query the user status and multicast program's real-time traffic through the CLI. The following table lists troubleshooting suggestions based on the query results.

NOTE

Data configurations are seldom changed in daily operation and maintenance. Therefore, multicast faults are usually caused by hardware problems. Hardware must be checked prior to data configuration during fault location.

 Copyright © Huawei Technologies Co., Ltd.

Table 13-42 Multicast emulation results and troubleshooting suggestions

Command	Results	Description	Troubleshooting Suggestions
display igmp user	The user status parameter **State** is **online**.	The multicast user can go online successfully.	• If the user can go online and the access device can communicate with the multicast source, the fault may be caused by a communication failure between the access device and the set-top box (STB). The reasons are as follows: – The hardware of the port on the access device is faulty. – The physical line between the access device and the modem is faulty. – The modem is faulty. – The STB is faulty. • If the user can go online but the traffic on the access device's uplink port is abnormal, the fault may be that the hardware connection between the access device and the upper-layer device (multicast router or multicast server) is incorrect, or the software configuration is incorrect. The common software configuration faults are as follows: – The remaining multicast bandwidth of the user is lower than the required

 Copyright © Huawei Technologies Co., Ltd.

Command	Results	Description	Troubleshooting Suggestions
			bandwidth of the ordered program. – The number of programs watched by the multicast user reaches the upper limit so that the user cannot order a new program. – The multicast user does not have the permission to watch the program. – The program ordered is not in the MVLAN to which the multicast user belongs. – The multicast user does not have the permission to order certain types of programs (such as HDTV). – The number of programs at a level watched by the multicast user reaches the upper limit so that the user cannot order a new program at this level. – The rate configured in the traffic profile bound to the traffic stream is far lower than the bandwidth of the multicast program. – There are too many prejoined static programs, occupying too many bandwidths.

 Copyright © Huawei Technologies Co., Ltd.

Command	Results	Description	Troubleshooting Suggestions
			The common software configuration problems of the upper layer device are as follows: – The program is not configured on the multicast server. – The TTL value set on the multicast server for the multicast stream is very small.

 Copyright © Huawei Technologies Co., Ltd.

Command	Results	Description	Troubleshooting Suggestions
	The user status parameter **State** is **offline**.	The multicast user fails to go online and therefore fails to order a program in the emulation.	The case is more complicated than the scenario in which the user is **online**. Handling suggestions are as follows: Enable the multicast debugging function and start the multicast emulation again to check whether the access device receives the report packet from the user for ordering a program. Run the following commands to enable the multicast debugging function: huawei(config)#**terminal debugging** huawei(config)#**terminal monitor** huawei(config)#**debugging igmp service-port** *index* **NOTE** **index** is the ID of the multicast user's service port. ● If the access device receives the report packet, the multicast link is normal but the access device fails to create a corresponding multicast entry. This is generally caused by incorrect multicast configurations on the access device. ● If the access device does not receive the report packet, the multicast link fails. This is mainly caused by incorrect access device data configurations, faulty physical link between the access device and the modem, or

 Copyright © Huawei Technologies Co., Ltd.

Command	Results	Description	Troubleshooting Suggestions
			hardware faults of terminals.
	The user status parameter **State** is **block**.	The multicast user is locked and therefore fails to order a program.	Run the **undo igmp user block** command to unblock the user.
display multicast flow-statistic	The multicast program's real-time traffic parameter **Multicast flow statistic result** is a small value or zero.	The program's traffic on the uplink port is small or zero.	The hardware connection between the access device and the upper-layer device (multicast router or multicast server) is incorrect, or the software configuration is incorrect. troubleshoot the fault based on the suggestions provided **when the user can go online but the traffic on the access device's uplink port is abnormal**.

 Copyright © Huawei Technologies Co., Ltd.

Command	Results	Description	Troubleshooting Suggestions
	The multicast program's real-time traffic parameter **Multicast flow statistic result** is close to the program's bandwidth.	The device uplink port can communicate with the multicast source.	• If the user can go online and the access device can communicate with the multicast source, the fault may be caused by a communication failure between the access device and the STB. The reasons are as follows: – The hardware of the port on the access device is faulty. – The physical line between the access device and the modem is faulty. – The modem is faulty. – The STB is faulty. • If the access device can communicate with the multicast source but the user cannot go online, troubleshoot the fault based on the suggestions provided **when the user status parameter State is offline**.

13.7.1.4 Configuration

Context

- The configuration of the multicast port is correct.
- The multicast user for whom the multicast emulation test is performed has the rights to watch the configured multicast programs.

Procedure

Step 1 Run the **igmp static-join** command to perform the multicast emulation test for the multicast user.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#btv
huawei(config-btv)#igmp static-join service-port 500
{ ip<K>|ipv6<K> }:ip
{ ip-addr<I><X.X.X.X> }:224.1.1.1
{ vlan<K> }:vlan
{ vlanid<U><1,4093> }:4002
```

Step 2 Run the **display igmp user** command to query the status of the multicast user.

- If the multicast user is in **offline** state, the multicast user fails to request for programs.
- If the multicast user is in **online** state, the multicast user requests for programs successfully.
- If the multicast user is in **block** state, the multicast user is blocked. In this case, run the **undo igmp user block** command to unblock the user.

```
huawei(config)#display igmp user service-port 500
  User                  : 0/1/1
State: online // The multicast user is online.
  Authentication          : auth
  Quick leave             : MAC-based
  IGMP flow ID            : 500
  Video flow ID           : 500
  Log switch              : enable
  Bind profiles           : 2
  IGMP version            : IGMP v3
  Current version         : IGMP v3
  ......
```

Step 3 Run the **display multicast flow-statistic** command to query the real-time traffic of the programs that the multicast user requests for in the multicast emulation test.

- If the real-time traffic of the multicast programs is a smaller value or 0, the multicast source does not deliver multicast programs or the multicast service stream does not arrive at the device. That is, the communication between the device and the multicast source is abnormal.
- If the real-time traffic of the multicast programs approaches the bandwidth of the multicast programs, the multicast source delivers the multicast programs to the device. That is, the communication between the device and the multicast source is normal.

```
huawei(config)#btv
huawei(config-btv)#display multicast flow-statistic vlan 4002 ip 224.1.1.1
  Command is being executed, please wait...
  Multicast flow statistic result: 29600(kbps)     //The real-time traffic of multicast program 224.1.1.1 is
29600 kbit/s. This indicates that the multicast source issues multicast traffic.
```

Step 4 Run the **undo igmp static-join** command to stop multicast emulation.

```
huawei(config)#btv
huawei(config-btv)#undo igmp static-join service-port 500
{ ip<K>|ipv6<K> }:ip
{ ip-addr<I><X.X.X.X> }:224.1.1.1
{ vlan<K> }:vlan
{ vlanid<U><1,4093> }:4002
```

----End

13.7.1.5 Multicast Emulation Reference Standards and Protocols

- RFC-2236: Fenner, W., "Internet Group Management Protocol, Version 2", RFC 2236, November 1997
- RFC 3376: B. Cain., "Internet Group Management Protocol, Version 3", RFC 3376, October 2002

 Copyright © Huawei Technologies Co., Ltd.

- RFC 4607: H. Holbrook, "Source-Specific Multicast for IP", RFC 4607, August 2006

13.7.2 Video Quality Monitoring

Video quality monitoring enables an access device to remotely monitor affected video programs on configured monitoring points for fault demarcation.

13.7.2.1 Introduction

Background

After a fault occurs in the video service of a residential user, the user reports this fault to the carrier. Then, the carrier's maintenance personnel detect and rectify the fault. Before rectifying the fault, the maintenance personnel must demarcate the fault, involving the following cases:

- If multiple users report the fault related to an IPTV or VoD program within a period of time, and these users connect to different optical line terminals (OLTs), the IPTV server or core switching network may be faulty. In this case, the maintenance personnel only need to troubleshoot the IPTV server or core switching network.
- If only few users report the fault, and these users connect to the same OLT, the maintenance personnel must diagnose the video quality on the network for fault demarcation.

However, the maintenance personnel relies on home visits for troubleshooting. In the fiber to the home (FTTH) scenario, optical network terminals (ONTs) are located at user homes. Therefore, onsite diagnosing is not only time-consuming but also costly. In this case, an effective method of remotely diagnosing video quality is urgently required.

Overview

In video quality monitoring, monitoring points are configured on the OLT and ONTs. Each monitoring point is used for monitoring quality indicators of each video program. Maintenance personnel can remotely enable video quality monitoring and obtain monitoring results on the NMS or OLT. Based on monitoring results obtained from each monitoring point, the maintenance personnel comprehensively identify fault points, thereby reducing fault locating costs.

Compared with the video quality monitoring enabled using the OLT, the video quality monitoring enabled using the NMS has the following advantages:

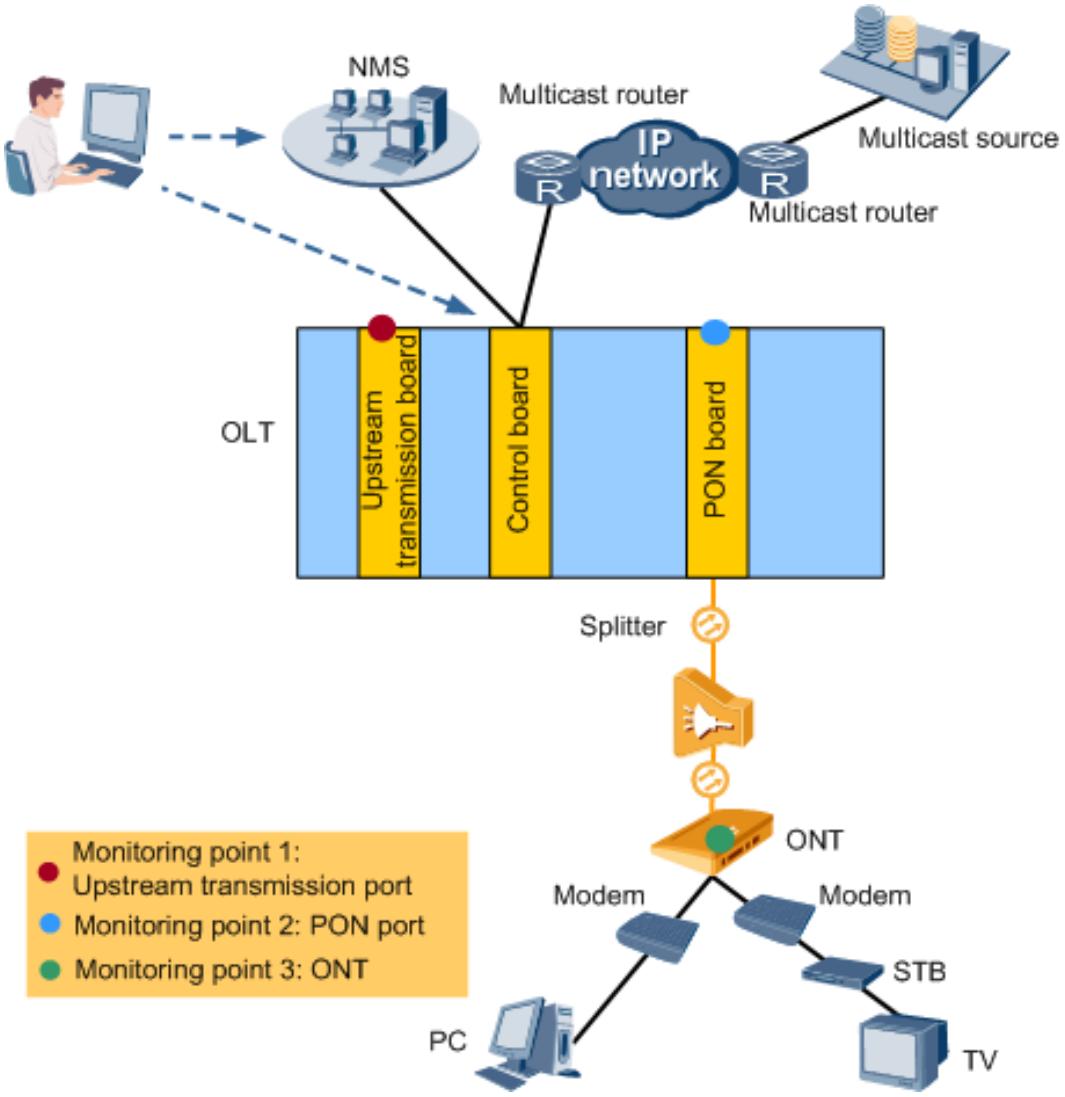
- The NMS can store more video quality monitoring results than the OLT. In addition, the NMS supports monitoring result dumping, meeting the requirement of long-term video quality monitoring. Major video quality-related faults occur randomly. Therefore, the identification of such faults requires long-term video quality monitoring data.
- The NMS can display monitoring results in graphs. These trend graphs for each channel of program facilitate information obtaining.

 Copyright © Huawei Technologies Co., Ltd.

Video quality monitoring points can be configured on uplink ports, PON ports of the OLT, and ONTs. In addition, the quality of both common videos and 4K videos can be monitored.

Figure 13-70 shows the application scenario of video quality monitoring.

Figure 13-70 Application scenario of video quality monitoring



13.7.2.2 Basic Concepts

Video quality monitoring involves the following quality indicators: media delivery index (MDI), mean opinion score for video (VMOS), and packet loss rate (PLR), which are described as follows:

MDI

MDI, defined in RFC 4445, is a quantitative number assigned to the quality of video streams transmitted over a network.

The MDI consists of two components: the delay factor (DF) and the media loss rate (MLR).

 Copyright © Huawei Technologies Co., Ltd.

- DF is the maximum difference, observed at the end of each media data packet, between the arrival of media data and the forwarding of media data. This indicator reflects the delay and jitter of the tested video stream. A large jitter leads to a large DF value, which promotes high requirements on the buffer of the decoding device. Therefore, the smaller the DF value, the better. Recommended DF value range is 0-50 ms. Any DF value less than 200 ms is accepted.
- MLR specifies the number of media data packets lost per second. This indicator determines video quality. By measuring MLR, maintenance personnel can detect, identify, and trace media data packet loss on a network. According to WT126, the maximum accepted MLRs for standard definition (SD)/VoD programs and high definition (HD) programs are 5 media data packets per 30 minutes and 5 media data packets per 240 minutes, respectively.

VMOS

VMOS is a subjective method of evaluating the quality of a video program.

A VMOS value is measured on a continuous scale of 1 to 5, representing the video quality of unsatisfactory, poor, fair, good, and excellent, respectively.

PLR

PLR specifies the number of IP packets lost per second. PLR is different from MLR, which specifies the number of media data packets lost in one sampling period. One IP packet contains seven media data packets.

By measuring PLR, maintenance personnel can detect, identify, and trace IP packet loss on a network.

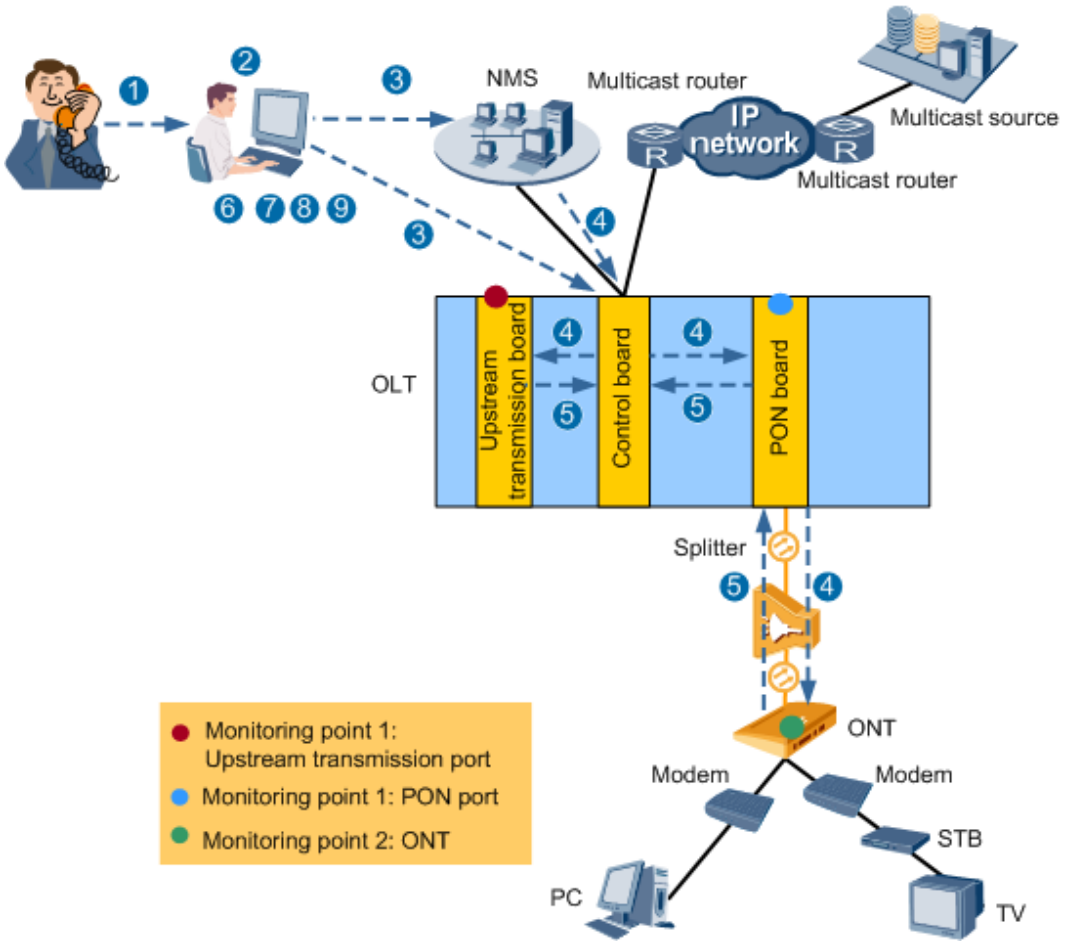
13.7.2.3 Principles

Implementation Process

Figure 13-71 shows the process of implementing video quality monitoring.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-71 Process of implementing video quality monitoring



The process is as follows:

1. After a fault occurs in the video service of a residential user, the user reports this fault to the carrier.
2. The carrier's maintenance personnel exclude communication faults after checking the head end system or performing an Internet Group Management Protocol (IGMP) emulation test.
3. The maintenance personnel enable video quality monitoring using the OLT or NMS.

   Video quality monitoring can be enabled on uplink ports, PON ports or ONTs.

   NOTE

   If video quality monitoring is enabled on an uplink port, only the quality of inbound video streams can be monitored. If video quality monitoring is enabled on a non-uplink port, only the quality of outbound video streams can be monitored.

   - To monitor the quality of an IPTV program, specify a multicast VLAN and multicast address.
   - To monitor the quality of a VoD program, specify the source and destination IP addresses of this program, destination IP address, destination port, and transmission protocol.
4. The OLT or NMS issues a command for enabling video quality monitoring to each monitoring point. After receiving this command, these monitoring points

 Copyright © Huawei Technologies Co., Ltd.

configure hardware based on the program information carried in the command. Then, the hardware captures video streams, adds timestamps to them, and sends the video streams to the CPU for monitoring indicator calculation.

5. The CPU sends calculated results to each monitoring point in real time, and the monitoring points periodically report these results to the control board of the OLT.
6. Video quality monitoring stops when the preset monitoring duration times out or the maintenance personnel disable this function.
7. The maintenance personnel obtain monitoring results from the OLT or NMS.
8. The maintenance personnel demarcate the fault based on monitoring results.
9. The maintenance personnel rectify the fault.

Fault Locating

After completing video quality monitoring, the maintenance personnel obtain monitoring results to learn program quality on each monitoring point for fault demarcation. **Table 13-43** shows fault demarcation rules.

Table 13-43 Fault demarcation rules

Monitoring Point on an Uplink Port	Monitoring Point on a PON Port	Monitoring Point on an ONT	Conclusion
POK	POK	POK	The fault may occur on the upper-layer network or head end system. Detect the upper-layer metropolitan area network (MAN) of the OLT.
OK	POK	POK	The fault may occur between the PON board and the control board/upstream transmission board on OLT. Detect the inner OLT.

 Copyright © Huawei Technologies Co., Ltd.

Monitoring Point on an Uplink Port	Monitoring Point on a PON Port	Monitoring Point on an ONT	Conclusion
OK	OK	POK	The fault may occur between the PON board and the outbound port on the ONT (including the inner ONT).
OK	OK	OK	The fault may occur at user home, such as the Internet network or set top box (STB).

13.7.2.4 Configuring Video Quality Monitoring

Prerequisites

The carrier's maintenance personnel have excluded communication faults after checking the head end system or performing an multicast emulation test.

Procedure

Step 1 (Optional) Run the **video-quality-monitor** command to configure video quality monitoring indicator thresholds and a monitoring period.

Perform this step if default video quality monitoring indicator thresholds and monitoring period fail to meet network requirements. The monitoring indicators include DF, MLR, VMOS, and PLR.

Step 2 Start video quality monitoring.

Run the **video-quality-monitor start** command to start a video quality monitoring instance. You can configure one monitoring point on an upstream port, GPON port or an ONT, or on all of them.

Step 3 (Optional) Stop video quality monitoring.

To stop video quality monitoring before the monitoring duration times out, run the **video-quality-monitor stop** command to stop the video quality monitoring instance.

Step 4 Query video quality monitoring results of the instance.

Run the **display video-quality-monitor result** command to query the monitoring results and demarcate the fault.

1. Query the 5-minute VMOS availability ratio (AR) on each monitoring point.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

The 5-minute VMOS AR is the ratio of the number of available VMOS seconds to the total duration of 300s (5 minute x 60 seconds). When the VMOS value is greater than or equal to the fair threshold configured in the VMOS threshold profile, the VMOS value is available.

2. If the VMOS AR of a monitoring point is low, further query the detailed results of the video quality monitoring instance for fault identification.

----End

Example

NOTE

Terminal display significantly varies based on video quality monitoring instances. Therefore, the following example does not provide terminal display for queried commands. For details about the terminal display and parameter description, see the OLT Command Reference.

The NMS provides monitoring results in graphs for simple information obtaining. Therefore, if the NMS is available, use it to enable video quality monitoring and query monitoring results.

The following is an example configuration of the video quality monitoring feature:

- Excellent, good, and fair VMOS thresholds are 44, 33, and 22, respectively.
- Video quality monitoring points are configured on upstream port 0/9/0, GPON port 0/2/0 and ONT 0 connected to port 0/2/0.
- The multicast VLAN is 100.
- The IP address of the multicast program is 224.1.1.1.

```
huawei(config)#video-quality-monitor vmos excellent-threshold 44 good-threshold 33 medium-threshold 22
huawei(config)#video-quality-monitor start port 0/9/0 port 0/2/0 ont 0/2/0 0 destination-ip 224.1.1.1 vlan 100
huawei(config)#display video-quality-monitor result port 0/9/0 ratio vmos
huawei(config)#display video-quality-monitor result port 0/2/0 ratio vmos
huawei(config)#display video-quality-monitor result ont 0/2/0 0 ratio vmos
huawei(config)#display video-quality-monitor result port 0/9/0 detail
huawei(config)#display video-quality-monitor result port 0/2/0 detail
huawei(config)#display video-quality-monitor result ont 0/2/0 0 detail
```

13.7.3 RTP Video Quality Monitoring

In Real-Time Transport Protocol (RTP) video quality monitoring, the quality of multicast video streams is monitored by means of RTP packet detection. Maintenance personnel can identify multicast service failures based on the detection results.

Basic Principle

RTP packets are encapsulated in User Datagram Protocol (UDP) packet headers, as shown in **Figure 13-72**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-72 IP+UDP packet format

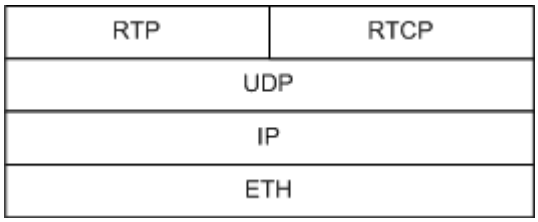
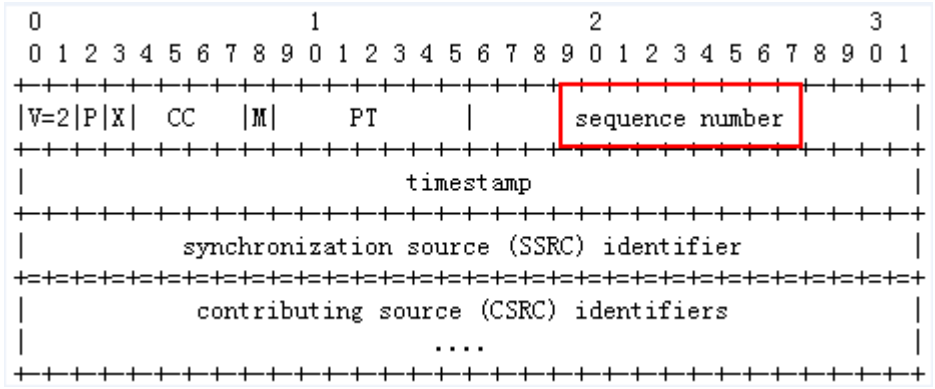


Figure 13-73 shows the format of the RTP packet header. **sequence number** indicates the sequence number of the RTP packet, with a length of 16 bits. The initial value is generated at random, and the sequence number is incremented by 1 when 1 RTP packet is sent. For details about the RTP packet header, see RFC3550.

Figure 13-73 Format of the RTP packet header



In RTP video quality monitoring, RTP sequence numbers in multicast program streams are periodically detected, to identify packet loss, error packets, and duplicate packets in the multicast program streams. During RTP packet loss detection, the system records the number of lost, error, and duplicate packets and records the start and end sequence numbers of packets within the detection period. The program segments with poor quality can be identified based on the start and end sequence numbers of packets within the detection period.

NOTE

- Statistics cannot be collected on segmented packets because segmented packets do not contain RTP packet headers. Therefore, it is recommended that the maximum transmission unit (MTU) in the system be set to less than 1500, to avoid network segmentation.
- The system clears all monitoring data upon active/standby switchover.
- For monitoring points in a link aggregation group, the command is delivered only to the specified port, without affecting other ports.
- For ports in a protection group, the command is delivered only to the active port. When port switchover is performed, the monitoring function must be configured again.

Operating Process

RTP video quality monitoring of the OLT applies to non-cascading and cascading scenarios. Figure 13-74 and Figure 13-75 show the RTP video quality monitoring processes in non-cascading and cascading scenarios respectively.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-74 RTP video quality monitoring process in non-cascading scenarios

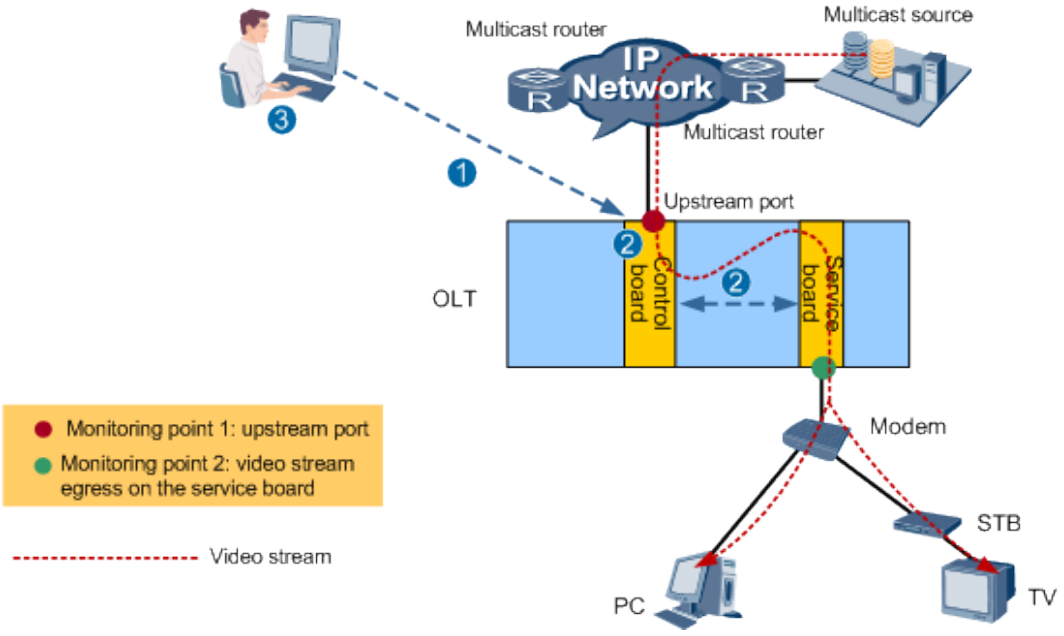
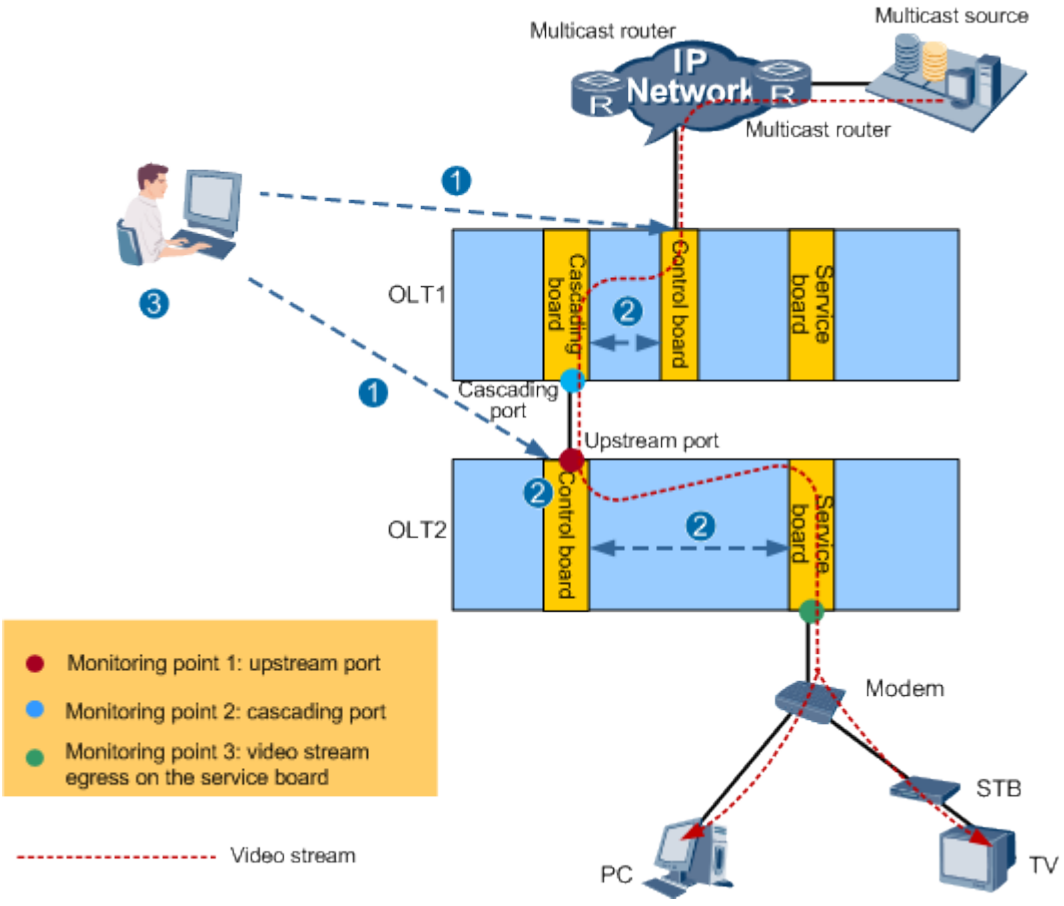


Figure 13-75 RTP video quality monitoring process in cascading scenarios



The RTP video quality monitoring process is as follows:

 Copyright © Huawei Technologies Co., Ltd.

1. Maintenance personnel run the **rtp multicast statistic** command on the OLT CLI to enable RTP video quality monitoring and specify program information. The OLT supports RTP video quality monitoring at upstream ports, cascading ports, and video stream egresses on service boards. Monitoring can be enabled for a single monitoring point or for all monitoring points in a unified manner.

   NOTE

   In cascading scenarios, RTP video quality monitoring can be enabled on OLT1 to monitor cascading ports, and can be enabled on OLT2 to monitor upstream ports or video stream egresses on service boards.

2. The control board of the OLT delivers an enabling message to the monitoring points. After receiving the enabling message, the monitoring points detect RTP packets based on the program information and periodically (*interval* configurable) send detection results to the CPU of the control board.
3. In the monitoring duration, maintenance personnel can run the **display rtp multicast statistic** command to view the monitoring results and perform troubleshooting based on the monitoring results.

   NOTE

   - The system records logs only for abnormal packets.
   - When the monitoring duration exceeds the preset value, the system clears all monitoring records.

Fault Diagnosis

After RTP video quality monitoring is completed, maintenance personnel query the monitoring results on the OLT CLI and obtain program quality of each monitoring point. Based on the monitoring results, maintenance personnel can perform fault diagnosis, as described in **Table 13-44**.

OK indicates that the total number of lost, duplicate, and error packets is 0. **POK** indicates that the total number of lost, duplicate, or error packets is not 0.

Table 13-44 Fault diagnosis based on monitoring results

Cascading Port	Upstream Port	Video Stream Ingress of a Service Board	Conclusion
POK	POK	POK	• The upper-layer access device (OLT1) is faulty. • The upper-layer network or headend system encounters a problem. Check the upper-layer MAN of the OLT.

 Copyright © Huawei Technologies Co., Ltd.

Cascading Port	Upstream Port	Video Stream Ingress of a Service Board	Conclusion
OK	POK	POK	The cascading devices are not properly connected.
OK	OK	POK	The lower-layer access device (OLT2) is faulty.
OK	OK	OK	The Internet connection fails, or the STB is faulty.

13.7.4 eMDI Video Quality Monitoring for the Multicast Service

Unlike video mean opinion score (VMOS), Media Delivery Index (MDI), or Real-Time Transport Protocol (RTP) video quality monitoring that features passive O&M, Enhanced Media Delivery Index (eMDI) video quality monitoring supports real-time multicast program monitoring and periodic result reporting to servers. It enables fast video fault locating, realizes proactive O&M, and guarantees the video services for VIP users.

13.7.4.1 Introduction

Context

For video quality monitoring, video program monitoring points are set on network devices to monitor the quality indicators of video programs. O&M personnel locate video faults by remotely querying and evaluating monitoring results, avoiding unnecessary door-to-door visits.

Existing VMOS, MDI, and RTP video quality monitoring solutions are all passive O&M solutions. For these solutions, carriers' O&M personnel enable video monitoring only after users report multicast artifacts or frame freezing. Such solutions have the following disadvantages:

- VMOS and MDI monitoring solutions feature high algorithm overheads, so long-time video monitoring in real time is not supported. After users report faults, O&M personnel can only wait for fault reproduction before locating.
- VMOS, MDI, and RTP monitoring solutions do not support large-scale concurrent program monitoring, so worldwide deployment is not supported.
- These solutions do not support monitoring indicators in forward error correction (FEC) and retransmission (RET) scenarios, so monitoring results in FEC and RET scenarios are inaccurate.

 Copyright © Huawei Technologies Co., Ltd.

eMDI video quality monitoring properly solves the preceding problems of VMOS, MDI, and RTP video monitoring solutions.

- For eMDI video quality monitoring, real-time monitoring can be enabled for a long time and monitoring results are uploaded to servers. O&M personnel can implement proactive O&M and guarantee the video services for VIP users by querying the monitoring results.
- An OLT supports concurrent eMDI video quality monitoring of 1024 programs. Therefore, the eMDI monitoring solution can be deployed on the entire network, covering numerous programs.
- eMDI video quality monitoring supports monitoring indicators in FEC and RET scenarios, so monitoring results are more convincing.

Table 13-45 compares 3 video quality monitoring solutions.

Table 13-45 Comparison of video quality monitoring solutions

Solution	Monitoring Indicator	Monitored Program Stream Type
VMOS/MDI video quality monitoring	• Video key quality indicators (KQIs) that reflect subjective user experience. • Video key performance indicators (KPIs) that reflect objective network conditions.	RTP-encapsulated and non-RTP-encapsulated video program streams
RTP video quality monitoring	Video KPIs	RTP-encapsulated MPEG2-TS video program streams
eMDI video quality monitoring	Video KPIs	RTP-encapsulated MPEG2-TS video program streams **NOTE** • For an OLT, eMDI, VMOS/MDI, and RTP video quality monitoring solutions are not mutually exclusive and can be enabled at the same time. • For an ONT, eMDI and VMOS/MDI video quality monitoring solutions cannot be enabled at the same time.

Overview

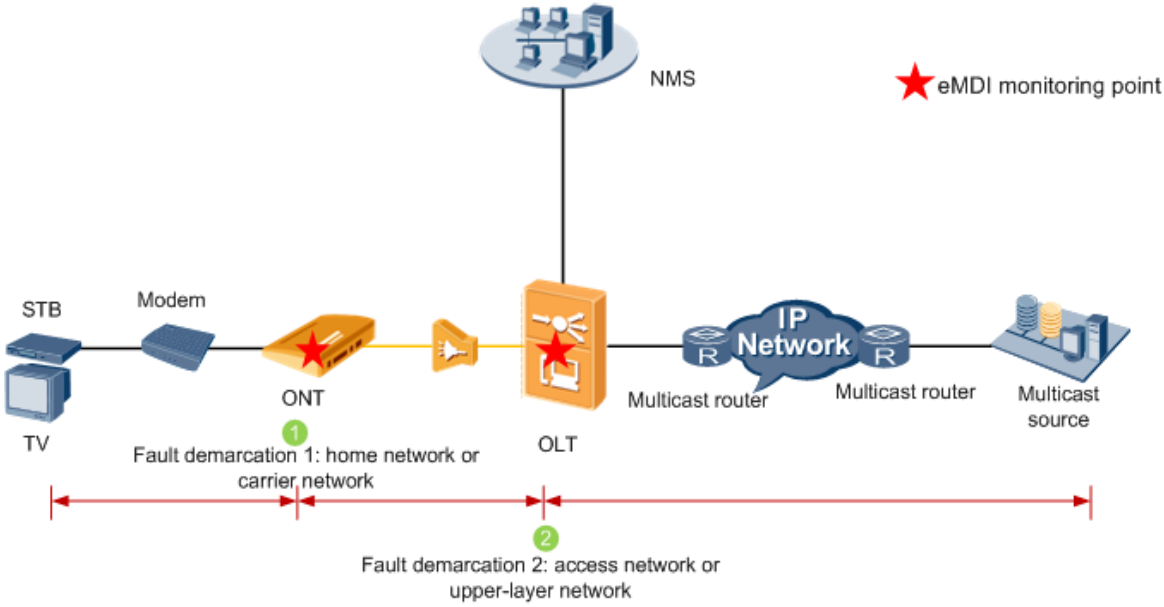
eMDI video quality monitoring is an enhanced solution on the basis of MDI video quality monitoring. The eMDI solution monitors videos based on the RTP sequence number, and calculates the RTP packet loss density and number of consecutively

 Copyright © Huawei Technologies Co., Ltd.

lost RTP packets, providing more accurate monitoring results than the MDI solution does in FEC and RET scenarios.

eMDI monitoring points can be set on an OLT board or ONT, as shown in **Figure 13-76**. O&M personnel can use the NMS or OLT CLI to enable and disable eMDI video quality monitoring, or query monitoring results.

Figure 13-76 eMDI video quality monitoring



Monitoring Point Location	Maximum Number of Concurrently Monitored Programs	Supported Encapsulation Format of Monitored Video Program Streams
ONT	1	• IPoE • PPPoE
OLT board	1024	IPoE

NOTE

- eMDI video quality monitoring supports only IPv4 video program streams and does not support IPv6 video program streams.
- eMDI video quality monitoring supports only live IPTV programs and does not support video on demand (VoD) programs.

13.7.4.2 eMDI Video Quality Monitoring Indicators

eMDI video quality monitoring indicators include RTP loss rate (RTP-LR), RTP sequence error (RTP-SE), RTP loss period (RTP-LP), and RTP effective loss factor for FEC (RTP-ELF).

Compared with MDI video quality monitoring, eMDI video quality monitoring supports a sequence number monitoring indicator based on the RTP packet

 Copyright © Huawei Technologies Co., Ltd.

header. During eMDI video quality monitoring, the RTP sequence number in a video program stream is periodically detected to determine the number of lost packets and that of disordered packets. **Figure 13-77** shows the format of the RTP packet header. Specifically, **sequence number** indicates the sequence number of an RTP packet and its length is 16 bits. Its initial value is random, and its value is increased by 1 with each RTP packet sent. For details about the RTP packet header, see RFC3550. The receiver detects the value of sequence number to determine whether packet loss occurs or disordered packets exist, as shown in **Figure 13-78**.

NOTE

Delay factor (DF) and media loss rate (MLR) indicators defined in MDI have great impacts on the CPU performance and long-time monitoring is not supported, so these 2 indicators are not measured for the OLT and its lower-level network devices during eMDI video quality monitoring.

Figure 13-77 Format of the RTP packet header

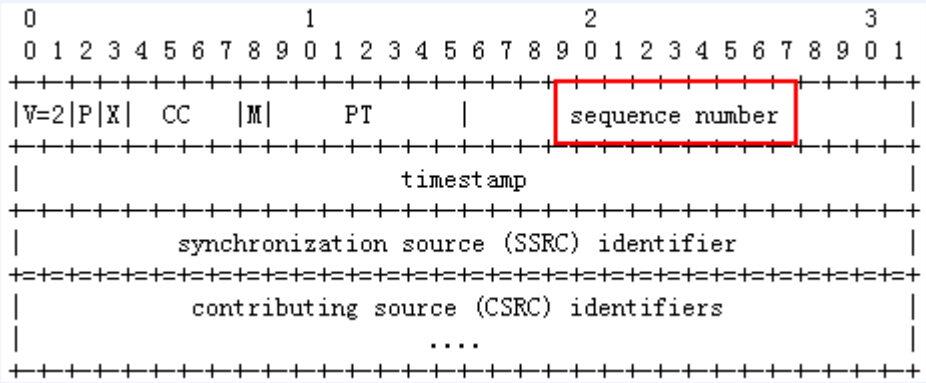


Figure 13-78 Example of detection based on the sequence number in the RTP packet header

	Packet Sequence Number	Detection Result
Normal	1.0 2.0 3.0 4.0	The packet sequence is normal. No packet is lost or disordered.
Packet loss	1.0 3.0 4.0	One packet is lost.
Disordered	1.0 3.0 2.0 4.0	One packet is disordered and one packet is lost.

RTP-LR

RTP-LR: indicates the number of lost packets in a statistical period. If this value is greater than 0, a video fault occurs.

RTP-SE

RTP-SE: indicates the number of disordered packets in a statistical period. If this value is greater than 0, a video fault occurs.

RTP-LP

RET is a packet loss recovery mechanism that introduces automatic retransmission requests when multicast or unicast packet loss occurs. This function reduces

 Copyright © Huawei Technologies Co., Ltd.

network quality deterioration caused by packet loss and improves the video quality of live and VoD programs.

RTP-LP: indicates the maximum number of consecutively lost RTP packets in a statistical period. If the RET threshold is set for a carrier's network, O&M personnel can compare RTP-LP and the threshold to determine whether a video fault occurs. When the RET function is enabled, a small number of lost packets will be retransmitted. A greater RTP-LP value indicates more difficult packet retransmission.

RTP-ELF

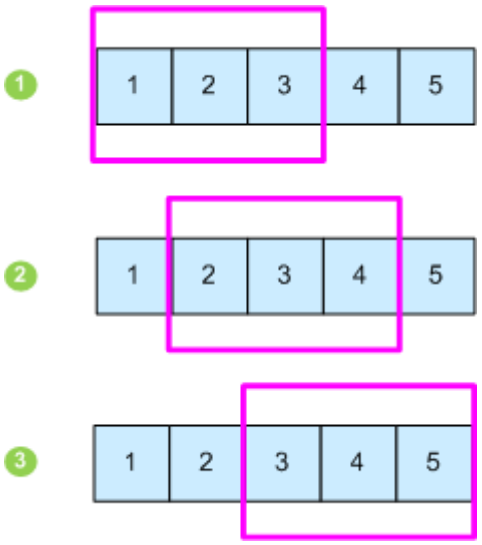
FEC is short for forward error correction. Redundant error-correcting codes can be added to transmission codes. When specified conditions are met, incorrect transmission codes can be automatically corrected during decoding, reducing the bit error rate of Rx signals and minimizing the frequency of artifacts and frame freezing.

RTP-ELF: indicates the effective loss factor for FEC in a statistical period. When the IPTV video service supports FEC and RET functions, a small number of data packets lost within a certain packet range or time period can be compensated through FEC and RET. This reduces the chance of IPTV video artifacts.

The value of RTP-ELF can be calculated by using the sliding window and threshold. In a statistical period, after a packet is slid, check whether the number of lost data packets in the sliding window is greater than the threshold. If the number of lost data packets is greater than the threshold, packet loss occurs. In this case, the number of packet loss windows is increased by 1, and window sliding stops when the monitoring period ends. If the total number of packet loss windows is greater than 0 in a statistical period, a video fault occurs.

As shown in **Figure 13-79**, if the sliding window contains 3 frames and the threshold is set to 2 frames, the sliding window can compensate 2 packets at most through FEC and RET. If more than 2 packets are lost, the excessive lost packets cannot be compensated, affecting video quality.

Figure 13-79 Sliding window



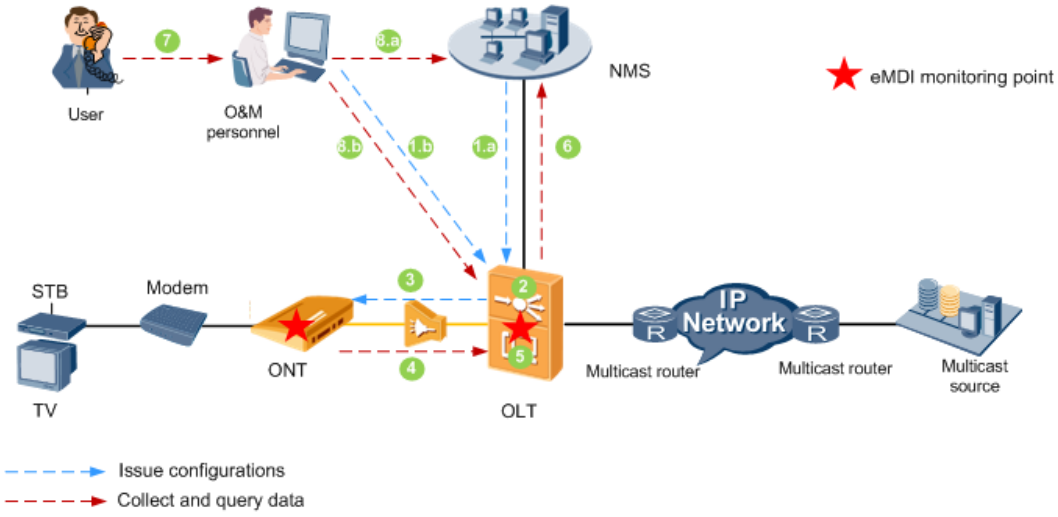
 Copyright © Huawei Technologies Co., Ltd.

13.7.4.3 Principle

Implementation Process

Figure 13-80 shows the implementation process of eMDI video quality monitoring.

Figure 13-80 Implementation process of eMDI video quality monitoring



The implementation process of eMDI video quality monitoring is as follows:

1. eMDI video quality monitoring is enabled.
   - O&M personnel can use the NMS to enable eMDI video quality monitoring of multicast programs on an OLT or ONT, and issue the enabling information to the OLT through MIB.
   - O&M personnel can run the **emdi monitor udp start** command to enable eMDI video quality monitoring of multicast programs on an OLT or ONT.

   NOTE

   - In scenarios where an inter-board aggregation group, protection group, or VRRP is configured on the multicast upstream port, if an inter-board switchover occurs on the ingress of monitored multicast program traffic, the eMDI video quality monitoring result reported during the switchover may be incorrect.
   - For an OLT, monitoring points can be set on the upstream board, cascade board, or service board. All multicast programs on a specified board or a specified program on a specified board can be monitored.
   - For an ONT, only one multicast program can be monitored at the same time. When enabling eMDI video quality monitoring, O&M personnel can specify an ONT and multicast program for monitoring, or specify an ONT without specifying any multicast program. If no multicast program is specified, when users order multiple programs, the system automatically selects the program with the highest data rate for monitoring. If monitoring is enabled with both the ONT and multicast program specified, after the specified multicast program is deleted, the system also selects the multicast program with the highest data rate for monitoring.
   - If the monitoring point is set on the upstream board of an OLT, only inbound video streams can be monitored; if the monitoring point is set on a non-upstream board of an OLT, only outbound video streams can be monitored.

 Copyright © Huawei Technologies Co., Ltd.

2. The OLT control board issues a message to enable monitoring of multicast programs of specified boards.
3. The OLT PON board records the eMDI monitoring configurations of the ONT and issues a message to the ONT. After receiving the message, the ONT monitors multicast programs in real time.
4. The OLT PON board collects the eMDI monitoring data of the ONT through OMCI or OAM messages every 5 minutes.
5. The OLT control board obtains the eMDI monitoring data of all boards every 5 minutes and saves the data.
6. The OLT control board reports monitoring results to the server every 5 minutes.

   eMDI fault data of the ONT and all OLT boards is summarized into a data file every 5 minutes. The OLT serves as an SFTP client and uploads this file to the SFTP server of the NMS after compressing the file. Rules for generating monitoring records are as follows:

   – When video traffic exists but no packet disorder or loss occurs, no monitoring record is generated, but the total number of packets needs to be recorded and reported.
   – When packet loss or disorder occurs, monitoring records are generated based on the monitoring period.
   – When no video traffic exists, no monitoring record is generated.

   The following table lists the parameters to be reported.

Parameter	Description
Time	Monitoring time, in DTS format.
Monitoring device	Device name or IP address
Monitoring point	• Subrack ID and slot ID • Subrack ID, slot ID, port ID, and ONT ID
Channel	Multicast program SIP, GIP, and VLAN ID
Data 1	OLT board data collected and summarized every 5 minutes **NOTE** The system generates a monitoring data record every minute. The OLT control board collects and summarizes the monitoring results of OLT boards with monitoring enabled, and reports the results to the server every 5 minutes. The OLT control board does not summarize the ONT monitoring data.

 Copyright © Huawei Technologies Co., Ltd.

Parameter	Description
Data 2	Monitoring data records with the maximum number of lost packets and disordered packets within 5 minutes **NOTE** The system generates a monitoring data record every minute. If the OLT control board reports all monitoring data to the server every 5 minutes, the service processing performance of the OLT will be affected. Therefore, the OLT control board uploads to the server a record with the maximum number of lost packets and another with the maximum number of disordered packets respectively for OLT boards and the ONT every 5 minutes.

7. A user reports a video program quality issue to the carrier.
8. O&M personnel remotely query historical monitoring data and locate the fault.
   - Query the records in all monitoring periods on the NMS.
   - Run the **display emdi monitor udp result** command on the OLT to query the monitoring records.

     NOTE

     Only the monitoring results of the current and previous statistical periods can be queried by running commands.
9. O&M personnel further troubleshoot the fault based on the fault locating result.

Fault Locating

After completing eMDI video quality monitoring, O&M personnel can query the monitoring results by running commands on the OLT or through the NMS, and learn the program quality of each monitoring point. Based on such information, they can locate the video quality fault by adhering to the following principles shown in **Table 13-46**.

 Copyright © Huawei Technologies Co., Ltd.

Table 13-46 Faulting locating roadmap for eMDI video quality monitoring

Monitoring Point on an OLT Upstream Board	Monitoring Point on an OLT Non-upstream Board	Monitoring Point on an ONT	Conclusion
POK	POK	POK	The upper-layer network or headend system may be faulty. O&M personnel should focus on checking the metropolitan area network (MAN) in the upstream direction of the OLT.
OK	POK	POK	The OLT may be faulty.
OK	OK	POK	The ONT or the line between the PON board and the ONT egress may be faulty.
OK	OK	OK	The interconnection or STB in the user's home may be abnormal.

13.7.4.4 Configuring eMDI Video Quality Monitoring

Prerequisites

The multicast service is configured.

Procedure

Step 1 (Optional) Run the **emdi interval** command to configure the period of eMDI video quality monitoring.

Perform this step if the default period of eMDI video quality monitoring does not meet actual network requirements. The default period is 60s.

Step 2 (Optional) Run the **emdi fec** command to configure the sliding window and threshold for eMDI video quality monitoring.

 Copyright © Huawei Technologies Co., Ltd.

In forward error correction (FEC) and retransmission (RET) scenarios, perform this step if the default sliding window and threshold for eMDI video quality monitoring do not meet actual network requirements. The default sliding window contains 100 frames and the threshold is 5 frames.

Step 3 Run the **emdi monitor udp start** command to enable an eMDI video quality monitoring instance.

- When eMDI video quality monitoring is enabled, if only the OLT board is specified, all programs on the board are monitored; if both the OLT board and program information (including the multicast VLAN, multicast IP address, and multicast source IP address) are specified, only the specified program on the specified board is monitored.
- When eMDI video quality monitoring is enabled, if only the ONT is specified, only the program with the highest data rate is monitored; if both the ONT and program information (including the multicast VLAN, multicast IP address, multicast source IP address, and destination port ID) are specified, only the specified program on the specified ONT is monitored.

Step 4 (Optional) To terminate video quality monitoring in advance, run the **emdi monitor udp stop** command to stop the created monitoring instance.

Step 5 Run the **file-server auto-backup emdi** command to configure the server to which eMDI monitoring data is uploaded.

Step 6 (Optional) Run the **display emdi monitor udp result** command to query the result of eMDI video quality monitoring, and perform fault locating based on the queried result.

Step 7 (Optional) Run the **display emdi monitor udp instance** command to query the information about the configured eMDI video quality monitoring instance.

Step 8 (Optional) Run the **display emdi info** command to check whether the period of eMDI video quality monitoring, sliding window, and threshold are correctly configured.

----End

Example

To configure the period of eMDI video quality monitoring to 30s, sliding window to 110 frames, and threshold to 10 frames, and enable the following 2 eMDI video quality monitoring instances, do as follows:

- Upstream port: 0/9, multicast VLAN: 100, multicast program IP address: 224.1.1.1.
- ONT port: GPON port 0/2/0, ONT ID: 0, multicast VLAN: 100, multicast program IP address: 224.1.1.2.

```
huawei(config)#emdi interval 30s
huawei(config)#emdi fec window 110 threshold 10
huawei(config)#emdi monitor udp start instance board 0/9 vlan 100 destination-ip 224.1.1.1
huawei(config)#emdi monitor udp start instance ont 0/2/0 0 vlan 100 destination-ip 224.1.1.2
huawei(config)#display emdi info
  ------------------------------------------------------------------------------
  eMDI interval (unit: s)                          : 30
  Window for RTP effective loss factor (unit: frame)       : 110
  Threshold for RTP effective loss factor (unit: frame)    : 10
  ------------------------------------------------------------------------------
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#display emdi monitor udp instance all
--------------------------------------------------------------------------------
Instance    F/S/P     ONT    VLAN     Destination IP      Source IP
--------------------------------------------------------------------------------
       0    0/9/-     -       100     224.1.1.1           -
       1    0/2/0     0       100     224.1.1.2           -
--------------------------------------------------------------------------------
Total: 2
```

13.7.5 eMDI Video Quality Monitoring for the Unicast Service

Enhanced Media Delivery Index (eMDI) video quality monitoring for the unicast service mainly applies to video services such as video surveillance, video of demand (VoD), and over the top (OTT) videos. It helps O&M personnel efficiently demarcate video quality faults in proactive O&M mode.

13.7.5.1 Introduction to eMDI for Unicast Services

Purpose of Introducing eMDI

As video services develop rapidly, users have higher requirements on video experience. Therefore, it is important to quickly demarcate and resolve video quality faults. Traditional demarcation requires O&M personnel to locate and analyze faults onsite, which is time-consuming and inefficient. Video quality monitoring can help O&M personnel remotely demarcate faults, avoiding unnecessary home visits and improving demarcation efficiency.

The existing Video Mean Opinion Score (VMOS), Media Delivery Index (MDI), and Real-Time Transport Protocol (RTP) video quality monitoring solutions can help O&M personnel remotely demarcate video quality faults. However, these solutions have the following limitations:

- **Passive O&M and low demarcation efficiency:** Video quality monitoring cannot be performed in real time for a long time. O&M personnel starts video quality monitoring only after a fault is reported, and can demarcate a fault only after it recurs.
- **Low demarcation accuracy**: These solutions do not support monitoring indicators in the forward error correction (FEC) or retransmission (RET) scenarios, resulting in inaccurate monitoring results in these scenarios.

  NOTE

  - RET is a lost packet recovery mechanism by automatic retransmission in multicast and unicast scenarios. This function can mitigate data quality deterioration caused by packet loss and improve the video quality of live TV and VOD programs.
  - FEC is a mechanism that adds redundant error correction codes to the transmission code column. When decoding is performed under certain conditions, incorrect codes can be automatically corrected. This reduces the bit error rate of received signals and the frequency of artifacts and frame freezing.
- **Low monitoring capacity**: These solutions can monitor only a small number of programs at the same time, and cannot be widely deployed on an entire network.
- **Limited services to monitor**: These solutions can demarcate only faults in video services carried by the User Datagram Protocol (UDP), and cannot demarcate faults in video services carried by the Transmission Control Protocol (TCP).

 Copyright © Huawei Technologies Co., Ltd.

To solve the preceding problems, Huawei launches the Enhanced Media Delivery Index (eMDI) video quality monitoring solution. eMDI offers the following benefits:

- **Proactive O&M and efficient demarcation**: The solution can monitor the video quality in real time over a long period of time and upload the monitoring results to a server. When a fault occurs, O&M personnel can quickly demarcate the fault based on the collected monitoring data.
- **Accurate demarcation**: For video services carried over UDP, effective packet loss factors are proposed under the FEC/RET compensation mechanism to accurately describe the impact of packet loss on video services and improve demarcation accuracy.
- **Network-wide deployment**: Multiple unicast video streams can be monitored at the same time. More video programs can be monitored through network-wide deployment.
- **More services to monitor**: The solution can monitor the quality of video services carried by both UDP and TCP.

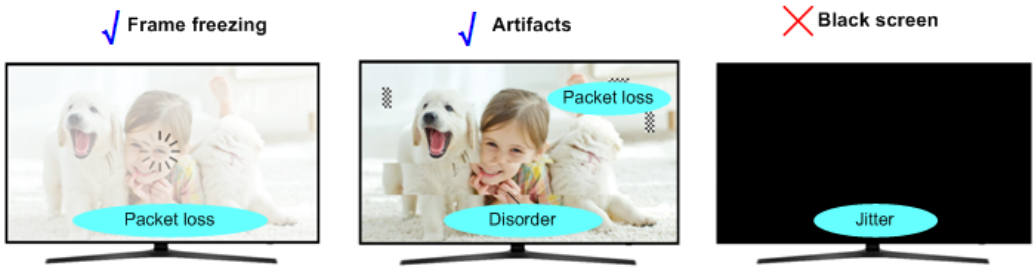
What is eMDI?

eMDI is an enhanced video quality monitoring solution developed based on MDI. It deploys monitoring points on network devices to monitor the quality of video streams in real time and periodically reports monitoring results to a server. By remotely viewing and evaluating monitoring results, O&M personnel can quickly demarcate video quality faults and provide assurance for VIP users.

Video fault symptoms to be monitored by eMDI

Packet loss, disorder, and jitter are three key factors that affect video quality. Even if the packet loss rate and disorder rate are less than 0.01%, frame freezing and artifacts can occur on a terminal, while jitter may cause black screens on a terminal. eMDI video quality monitoring applies only to frame freezing and artifact faults in videos.

Figure 13-81 Video fault symptoms that can be monitored by eMDI



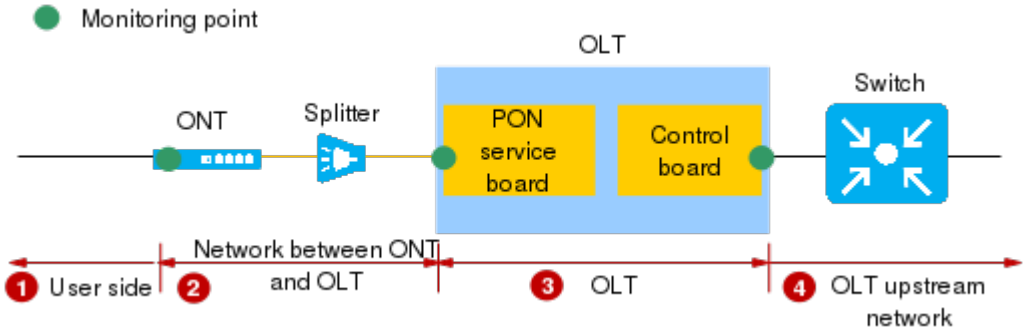
Positions for deploying monitoring points

eMDI video quality monitoring for unicast services supports monitoring of upstream and downstream unicast UDP/TCP packets. eMDI monitoring points can be deployed on the ONT, PON service port of the OLT, and upstream port of the OLT. The three monitoring points can be used to quickly demarcate faults on four network segments, as shown in **Figure 13-82**.

 Copyright © Huawei Technologies Co., Ltd.

O&M personnel can also deploy monitoring points at one or two positions. For example, if monitoring points are deployed only on PON service ports of the OLT, it can be determined that a fault occurs in the OLT or OLT upstream network when packet loss or disorder occurs in downstream video streams of a PON service port. When packet loss or disorder is detected in the upstream video stream of a PON service port, it can be determined that a fault occurs in the downstream network and devices of the OLT.

Figure 13-82 eMDI video quality monitoring points



Fault Locating

After eMDI video quality monitoring is enabled, O&M personnel can query the monitoring results on the OLT command line interface (CLI) or on the server to learn about the program quality on each monitoring point for fault demarcation. The following table shows fault demarcation rules.

Table 13-47 Fault demarcation based on eMDI video quality monitoring for downstream video streams (POK indicates that packet loss or disorder occurs, and OK indicates that no packet loss or disorder occurs.)

Monitoring Point on OLT Upstream Port	Monitoring Point on OLT PON Service Port	Monitoring Point on ONT	Conclusion
POK	POK	POK	The fault may occur in the upper-layer network or headend system. Check the metro network above the OLT.
OK	POK	POK	The fault may occur inside the OLT.
OK	OK	POK	The ONT or the line between the PON board and the ONT egress port may be faulty.
OK	OK	OK	The fault may occur in the user's home network (Internet or STB).

 Copyright © Huawei Technologies Co., Ltd.

Table 13-48 Fault demarcation based on eMDI video quality monitoring for upstream video streams (POK indicates that packet loss or disorder occurs, and OK indicates that no packet loss or disorder occurs.)

Monitoring Point on OLT Upstream Port	Monitoring Point on OLT PON Service Port	Monitoring Point on ONT	Conclusion
POK	POK	POK	The fault may occur in the user's home network (Internet or STB).
POK	POK	OK	The ONT or the line between the PON board and the ONT egress port may be faulty.
POK	OK	OK	The fault may occur inside the OLT.
OK	OK	OK	The fault may occur in the upper-layer network or headend system. Check the metro network above the OLT.

13.7.5.2 eMDI Monitoring Indicators for Unicast Services

eMDI video quality monitoring periodically collects monitoring data at monitoring points and uploads the data to the server for O&M personnel to analyze and demarcate faults. The eMDI monitoring indicators for UDP and TCP services are different.

Indicators for monitoring UDP video services

For video services carried over UDP, faults are analyzed by monitoring the sequence numbers of Real-time Transport Protocol (RTP) packets in video program streams. RTP is a transport layer protocol that provides real-time end-to-end data transmission services. It is widely used in communication and entertainment over streaming media, such as webcast, network TV, and video conference.

Figure 13-83 shows the format of an RTP packet header. **sequence number** indicates the sequence number of an RTP packet. The initial value is randomly generated and is incremented by 1 each time an RTP packet is sent. By monitoring the sequence numbers of RTP packets on monitoring points, O&M personnel can determine whether packet loss or disorder occurs in the video program stream, as shown in **Figure 13-84**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-83 RTP packet format

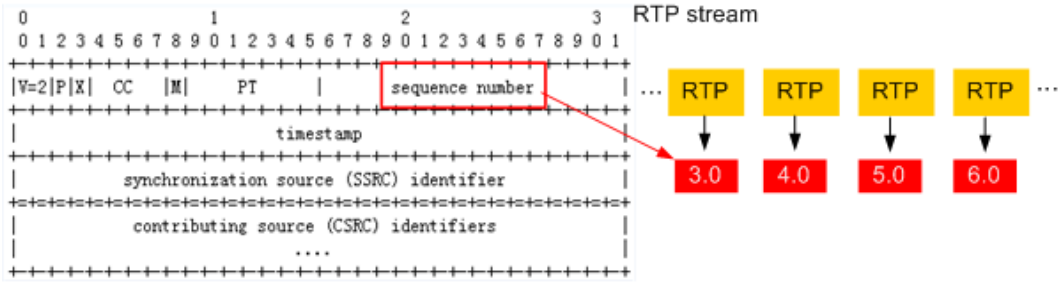


Figure 13-84 Packet monitoring based on sequence numbers in RTP packet headers

Table 13-49 lists the monitoring indicators for UDP video streams

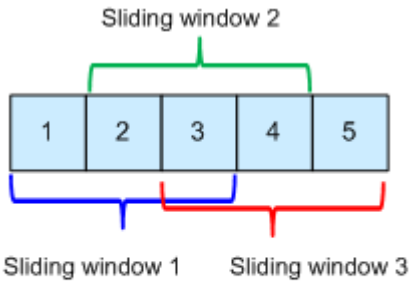
Table 13-49 Monitoring indicators for UDP video streams

Monitoring Indicator	Definition	Formula
RTP-LC	RTP Loss Count: number of lost packets in a statistical period	If the value of RTP-LC is greater than 0, a fault has occurred.
RTP-LR	RTP Loss Rate: rate of lost packets in a statistical period	RTP-LR = Number of lost packets/ (Number of received packets + Number of lost packets - Number of disordered packets) If the value of RTP-LP is greater than 0, a fault has occurred.
RTP-SE	RTP Sequence Error: number of disordered RTP packets in a statistical period	If the value of RTP-SE is greater than 0, a fault has occurred.

 Copyright © Huawei Technologies Co., Ltd.

Monitoring Indicator	Definition	Formula
RTP-SER	RTP Sequence Error Rate: rate of disordered RTP packets in a statistical period	RTP-SER = Number of disordered packets/(Number of received packets + Number of lost packets - Number of disordered packets) If the value of RTP-SER is greater than 0, a fault has occurred.
RTP-LP	RTP Loss Period: maximum number of consecutively lost RTP packets in a statistical period	If the RET threshold is set for an operator's network, O&M personnel can compare RTP-LP and the threshold to determine whether a fault has occurred. If RTP-LP is greater than the threshold, a fault has occurred.

 Copyright © Huawei Technologies Co., Ltd.

Monitoring Indicator	Definition	Formula
RTP-ELF	RTP Effective Loss Factor for FEC: effective FEC packet loss factor in a statistical period	If the IPTV video service supports FEC, a small number of data packets lost within a certain packet range or time period can be compensated through FEC. This reduces the probability of IPTV video artifacts. RTP-ELF can be calculated based on the number of packets in a sliding window (L) and the packet loss threshold (R). In a statistical period, each time a packet is sent, the system checks whether the number of lost packets in the sliding window L is greater than the packet loss threshold R. If yes, the window is a valid packet loss window. In this case, the value of RTP-ELF increases by 1 until the monitoring period ends, as shown in **Figure 13-85**. If the value of RTP-ELF is greater than 0, a fault has occurred. **Figure 13-85** Sliding window  Note: If the number of lost packets that cannot be compensated within the sliding window is greater than or equal to 1, the window is an effective packet loss window.

Indicators for monitoring TCP video services

Faults in TCP video services are analyzed based on TCP packet monitoring results. TCP is a connection-oriented, reliable, and byte stream-based transport layer communication protocol. **Table 13-50** lists the monitoring indicators for TCP video streams. O&M personnel compare the TCP monitoring indicators with the thresholds set on a server to demarcate faults.

 Copyright © Huawei Technologies Co., Ltd.

Table 13-50 Monitoring indicators for TCP video streams

Monitoring Indicator	Definition	Formula
MFR	Mean Flow Rate: average bitrate in a statistical period	MFR = Total length of packets received within the statistical period/Actual valid streaming time (unit: kbit/s) If the value of MFR is small, the server transmits data for a limited period of time, and no data is transmitted for an extended period during video streaming.
UPLR	Upstream Packet Loss Rate: packet loss rate on the upstream devices of a monitoring point within a statistical period	If no packet is lost, the sequence number of the current packet plus the packet length is equal to the expected sequence number of the next packet. If the sequence number of a packet is greater than expected, packet loss has occurred on an upstream device. The number of lost packets can be calculated based on the average packet size. UPLR = Number of upstream lost packets/(Total number of received packets + Total number of lost packets) UPLR can be used to determine whether packet loss occurs in the upstream of a monitoring point.

 Copyright © Huawei Technologies Co., Ltd.

Monitoring Indicator	Definition	Formula
DPLR	Downstream Packet Loss Rate: packet loss rate on the downstream devices of a monitoring point within a statistical period	If no packet is lost, the sequence number of the current packet plus the packet length is equal to the expected sequence number of the next packet. If the sequence number of a packet is smaller than expected, the packet is considered as a retransmitted packet. The number of retransmitted packets is considered as the total number of lost packets. Number of downstream lost packets = Total number of lost packets - Number of upstream lost packets DPLR = Number of downstream lost packets/(Total number of received packets + Total number of lost packets) DPLR can be used to determine whether packet loss occurs in the downstream of a monitoring point.
DRTT	Downstream Round-Trip Time (DRTT): average downstream two-way latency at a monitoring point in a statistical period	DRTT = T2 – T1 A received non-retransmitted packet is randomly selected, and its timestamp is recorded as T1. The expected sequence number of the next packet is calculated based on the sequence number and length of the current packet. If the sequence number of an upstream packet sent from a downstream device is greater than or equal to the expected sequence number, the timestamp of this packet is recorded as T2.

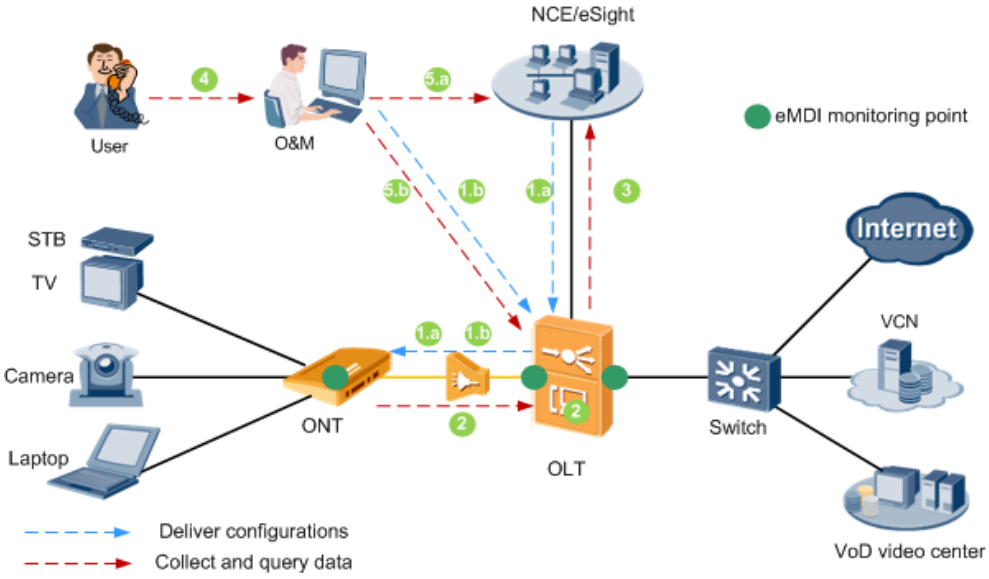
13.7.5.3 eMDI Interaction Process for Unicast Services

eMDI video quality monitoring for unicast services can be deployed in the upstream and downstream directions for both UDP and TCP video streams.

Figure 13-86 shows the interaction process of eMDI video quality monitoring.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-86 Interaction process of eMDI video quality monitoring



1. O&M personnel deploy monitoring points as required and start eMDI video quality monitoring for the OLT and ONTs on the NCE/eSight server or CLI.

   NOTE

   For unicast packets, the video stream to be monitored needs to be specified using a quintuple which includes the protocol type (TCP/UDP), destination IP address, source IP address, destination port number, and source port number.
2. If there is a monitoring point on an ONT, the ONT periodically reports video monitoring data to the OLT.
3. The OLT summarizes the video monitoring data on the ONT and OLT monitoring points and periodically reports the data to the NCE/eSight server.
4. After a fault occurs in the video service of a user, the user reports this fault to the operator.
5. O&M personnel can remotely view historical monitoring data on the NCE/eSight server or by running commands, and evaluate and demarcate the fault based on the historical monitoring data.
6. O&M personnel can troubleshoot the fault based on the demarcation result.

13.7.5.4 Network Application of eMDI for Unicast Services

Unicast video services include video backhaul, VOD, and OTT video services. eMDI video quality monitoring helps O&M personnel efficiently demarcate faults in such services.

eMDI Video Quality Monitoring in the Video Backhaul Scenario

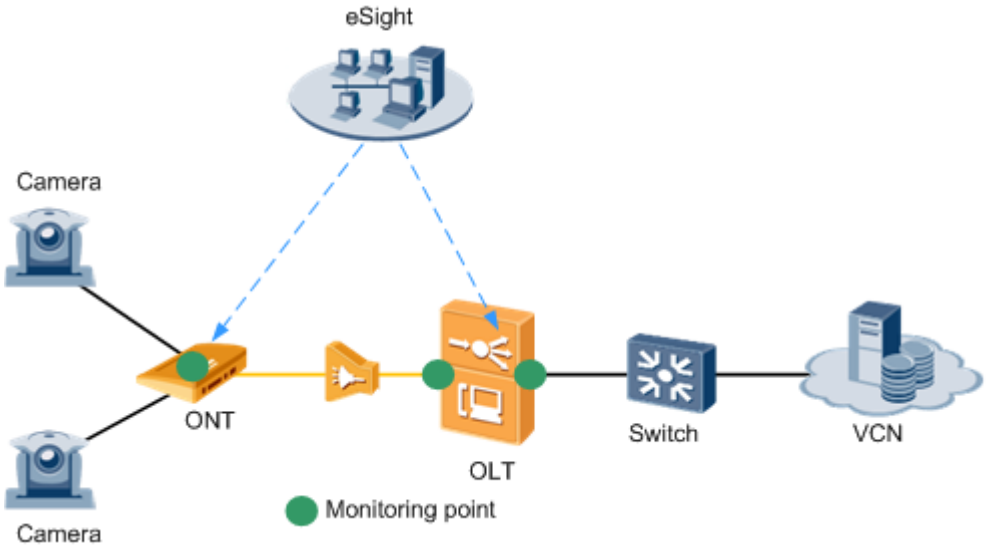
Regarding the O&M solution for a video backhaul network, customers are concerned about the online rate and intact rate of IP cameras (IPCs). In complex networks with multiple systems, if IPCs go offline or experience faults such as packet loss, the fault may be difficult to demarcate and troubleshoot, resulting in high O&M costs and hindering the development of the public security video

 Copyright © Huawei Technologies Co., Ltd.

backhaul networks. eMDI video quality monitoring can help O&M personnel resolve this problem.

On a video backhaul network shown in **Figure 13-87**, O&M personnel can use the eSight to deploy eMDI monitoring points on OLT upstream ports, OLT PON access ports, and ONTs to monitor video streams and collect monitoring indicators. When a video cloud node (VCN) detects that the data transmission quality deteriorates, O&M personnel can demarcate the fault by analyzing collected monitoring indicators.

Figure 13-87 eMDI video quality monitoring in the video backhaul scenario



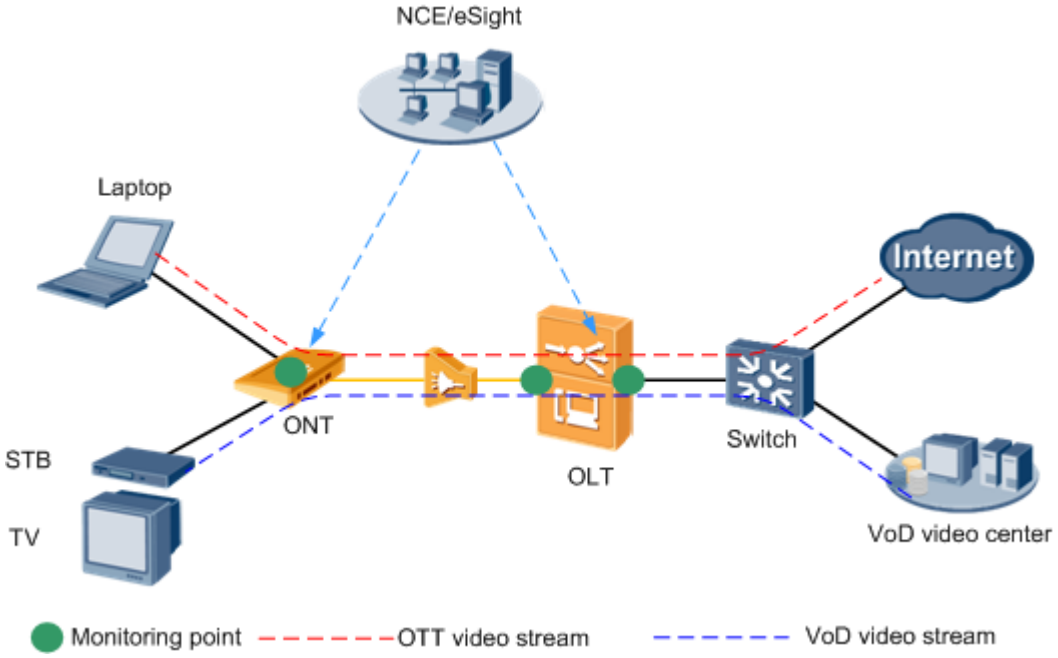
eMDI Video Quality Monitoring for VOD and OTT Scenarios

With the emergence of video services, users have ever higher requirements on video quality, and video quality monitoring becomes very important. eMDI video quality monitoring helps O&M personnel remotely demarcate faults, avoiding unnecessary home visits.

As shown in **Figure 13-88**, O&M personnel can use the NCE/eSight server to deploy eMDI monitoring points on OLT upstream ports, OLT PON access ports, and ONTs to monitor video streams and collect monitoring indicators. After a user reports a fault, O&M personnel can demarcate the fault by analyzing collected monitoring indicators.

 Copyright © Huawei Technologies Co., Ltd.

Figure 13-88 eMDI video quality monitoring for VOD and OTT video scenarios



13.7.5.5 Configuring eMDI Video Quality Monitoring

This topic describes the process of configuring eMDI video quality monitoring for unicast services.

Prerequisites

A video service has been configured.

Context

Table 13-51 Default settings of eMDI video quality monitoring

Parameter	Default Value
Monitoring Period	60s
Sliding Window	100 frames
FEC Threshold	5 frames

Procedure

Step 1 Run the **file-server auto-backup emdi-unicast** command to configure the server to which eMDI monitoring data is uploaded.

Step 2 (Optional) Run the **emdi interval** command to configure the period of eMDI video quality monitoring.

Perform this step if the default period of eMDI video quality monitoring does not meet actual network requirements.

 Copyright © Huawei Technologies Co., Ltd.

Step 3 (Optional) Run the **emdi fec** command to configure the sliding window and FEC threshold for eMDI video quality monitoring.

Perform this step if FEC and RET are enabled but the default sliding window and FEC threshold for eMDI video quality monitoring do not meet actual network requirements.

Step 4 Start an eMDI video quality monitoring instance.

- For UDP packets, run the **emdi monitor udp start** command.
- For TCP packets, run the **emdi monitor tcp start** command.

When starting eMDI video quality monitoring for a unicast service, you need to specify the monitoring point, monitoring direction (inbound: downstream; outbound: upstream), and source and destination IP addresses and port numbers of monitored video streams.

Step 5 (Optional) To cancel video quality monitoring in advance, run the either of the following commands:

- For UDP packets, run the **emdi monitor udp stop** command.
- For TCP packets, run the **emdi monitor tcp stop** command.

Step 6 (Optional) Query the eMDI video quality monitoring result and demarcate faults based on the query result.

- For UDP packets, run the **display emdi monitor udp result** command.
- For TCP packets, run the **display emdi monitor tcp result** command.

Step 7 (Optional) Query information about a configured eMDI video quality monitoring instance.

- For UDP packets, run the **display emdi monitor udp instance** command.
- For TCP packets, run the **display emdi monitor tcp instance** command.

Step 8 (Optional) Run the **display emdi info** command to check whether the period of eMDI video quality monitoring, sliding window, and threshold are correctly configured.

----End

Example

Example 1: Configure eMDI video quality monitoring for unicast UDP packets, with the data upload server IP address of 10.20.20.1, monitoring period of 30s, sliding window of 110 frames, and FEC threshold of 10 frames. Start the following three eMDI video quality monitoring instances:

- On upstream port 0/9/0, monitor the inbound unicast video stream whose destination IP address is 10.10.10.1, source IP address is 10.10.10.2, destination port number is 30, source port number is 30; and report the monitoring data to the server in real time.
- On PON port 0/2/0, monitor the inbound unicast video stream whose destination IP address is 10.10.10.1, source IP address is 10.10.10.2, destination port number is 30, source port number is 30; and report the monitoring data to the server in real time.
- On ONT 0 connected to GPON port 0/2/0, monitor the inbound unicast video stream whose destination IP address is 10.10.10.1, source IP address is

 Copyright © Huawei Technologies Co., Ltd.

10.10.10.2, destination port number is 30, source port number is 30; and report the monitoring data to the server in real time.

```
huawei(config)#file-server auto-backup emdi-unicast primary 10.20.20.1 sftp user
huawei(config)#emdi interval 30s
huawei(config)#emdi fec window 110 threshold 10
huawei(config)#emdi monitor udp start instance port 0/9/0 inbound destination-ip 10.10.10.1 source-ip 10.10.10.2 destination-port 30 source-port 30 instance-type realtime
huawei(config)#emdi monitor udp start instance port 0/2/0 inbound destination-ip 10.10.10.1 source-ip 10.10.10.2 destination-port 30 source-port 30 instance-type realtime
huawei(config)#emdi monitor udp start instance ont 0/2/0 0 inbound destination-ip 10.10.10.1 source-ip 10.10.10.2 destination-port 30 source-port 30 instance-type realtime
huawei(config)#display emdi info
  ------------------------------------------------------------------------------
  eMDI interval (unit: s)                          : 30
  Window for RTP effective loss factor (unit: frame)      : 110
  Threshold for RTP effective loss factor (unit: frame)   : 10
  ------------------------------------------------------------------------------
huawei(config)#display emdi monitor udp instance all unicast
  ------------------------------------------------------------------------------
  Instance    F/S/P     ONT    VLAN     Destination IP      Source IP          Instance
                                        Destination port    Source port    Dir   Type
  ------------------------------------------------------------------------------
       0    0/9/0     -        -     10.10.10.1          10.10.10.2
                                      30                  30             IN  realtime
       1    0/2/0     -        -     10.10.10.1          10.10.10.2
                                      30                  30             IN  realtime
       2    0/2/0     0        -     10.10.10.1          10.10.10.2
                                      30                  30             IN  realtime
  ------------------------------------------------------------------------------
  Total: 3
```

Example 2: Configure eMDI video quality monitoring for unicast TCP packets, with the data upload server IP address of 10.20.20.2, monitoring period of 30s, sliding window of 110 frames, and threshold of 10 frames. Start the following three eMDI video quality monitoring instances:

- On upstream port 0/9/0, monitor the outbound unicast video stream whose destination IP address is 10.10.10.3, source IP address is 10.10.10.4, destination port number is 40, source port number is 40; and report the monitoring data to the server periodically.
- On PON port 0/2/0, monitor the outbound unicast video stream whose destination IP address is 10.10.10.3, source IP address is 10.10.10.4, destination port number is 40, source port number is 40; and report the monitoring data to the server periodically.
- On ONT 0 connected to GPON port 0/2/0, monitor the outbound unicast video stream whose destination IP address is 10.10.10.3, source IP address is 10.10.10.4, destination port number is 40, source port number is 40; and report the monitoring data to the server periodically.

```
huawei(config)#file-server auto-backup emdi-unicast primary 10.20.20.2 sftp user
huawei(config)#emdi interval 30s
huawei(config)#emdi fec window 110 threshold 10
huawei(config)#emdi monitor tcp start instance port 0/9/0 outbound destination-ip 10.10.10.3 source-ip 10.10.10.4 destination-port 40 source-port 40 instance-type cycle
huawei(config)#emdi monitor tcp start instance port 0/2/0 outbound destination-ip 10.10.10.3 source-ip 10.10.10.4 destination-port 40 source-port 40 instance-type cycle
huawei(config)#emdi monitor tcp start instance ont 0/2/0  0 outbound destination-ip 10.10.10.3 source-ip 10.10.10.4 destination-port 40 source-port 40 instance-type cycle
huawei(config)#display emdi info
  ------------------------------------------------------------------------------
  eMDI interval (unit: s)                          : 30
  Window for RTP effective loss factor (unit: frame)      : 110
  Threshold for RTP effective loss factor (unit: frame)   : 10
  ------------------------------------------------------------------------------
huawei(config)#display emdi monitor tcp instance all
```

 Copyright © Huawei Technologies Co., Ltd.

```
-------------------------------------------------------------------------------
Instance    F/S/P     ONT    VLAN     Destination IP      Source IP          Instance
                                 Destination port    Source port    Dir   Type
-------------------------------------------------------------------------------
       0    0/9/0     -        -     10.10.10.3         10.10.10.4
                                 40                30           OUT   cycle
       1    0/2/0     -        -     10.10.10.3         10.10.10.4
                                 30                30           OUT   cycle
       2    0/2/0     0        -     10.10.10.3         10.10.10.4
                                 30                30           OUT   cycle

-------------------------------------------------------------------------------
Total: 3
```

13.7.6 Common Multicast Maintenance Methods

This topic describes only the principles of fault diagnosis for multicast services. For details about troubleshooting (from fault symptom to troubleshooting procedure), see the *Troubleshooting* manual.

User Log

The device log records the program order history of users. The log includes the port to which a user is connected, IP address of the program group, multicast VLAN (MVLAN), time when a user starts watching a program, time when a user stops watching a program, and log mode (for example, watch, preview, idle, or preview threshold crossing).

A log is generated in any of the following scenarios:

- Normal channel switching. In such switching, the interval between receiving a leave packet to receiving a join packet is longer than the defined time.
- Ordering failure. The common causes of an ordering failure are as follows:
  - The user does not have the right to watch the program.
  - The maximum number of programs the user can watch concurrently is exceeded.
  - The bandwidth CAC fails.
- Daily preview threshold crossing.
- Quiet leaving. In quiet leaving, the user does not respond to the general query of the device.
- Long-time program watching. "Long-time watching" means that the watch time reaches the maximum duration configured in the system.
- Operations that cause a user to go offline, for example, deleting or blocking a user.

Logs can be queried according to different query criteria, including by user, by program, by a specified period with regard to a user, and by a specified period with regard to a program. If users need to learn only the log quantity, the log statistics function is recommended. This frees users from reading multiple pages of numerous logs that are generated after the device has been running for a long time.

Users can also use the log clearing function to delete unwanted old logs. Logs of all users or of a specified user be deleted.

 Copyright © Huawei Technologies Co., Ltd.

Table 13-52 Commonly used multicast log commands

Command	Function
display igmp log	Query multicast logs.
display multicast failure-log	Query the multicast online and offline failure logs.
igmp log reset	Manually delete multicast logs.

IGMP Packet Statistics

A multicast program is ordered using the Internet Group Management Protocol (IGMP). Therefore, correctly sending, receiving, and processing IGMP packets is the prerequisite for successful program ordering. To facilitate fault locating in IGMP packet transmission/reception, the device supports three levels of IGMP packet statistics: global level, MVLAN level, and traffic stream level.

On the network side, the number of received IGMP query packets and number of sent IGMPv2/v3 join/leave packets can be queried based on MVLAN. According to the packet count, whether the upper-layer router is faulty can be determined.

On the user side, the number of received IGMPv2/v3 join/leave packets and number of sent IGMP query packets can be queried based on traffic stream. According to the packet count, whether a device in the home network is faulty can be determined.

Run the **display igmp statistic** command to query the multicast packet statistics.

Multicast Traffic Statistics Query

By querying multicast traffic statistics, users can determine whether multicast data reaches the ingress/egress of the device at the forwarding layer or whether packet loss occurs due to a low rate.

The device supports four query modes on the network side:

- A1: querying the number of sent/received multicast packets on an Ethernet port. Run the **display port statistics** command to query the statistics. The command output contains **Number of transmitted multicast frames** and **Number of received multicast frames**.
- A2: querying the ingress traffic (unit: kbit/s) of a specified multicast program (a pre-configured or dynamic program) or of a specified multicast upstream port. Run the **display multicast flow-statistic** command to query the traffic.
- A3: querying the number of required multicast packets (filtered by ACL) in the inbound direction of an Ethernet port.
- A4: querying the number of sent/received multicast frames on a PON port of an MDU that uses GPON upstream transmission. Run the **display gpon-port statistic** command to query the number. The command output contains **Received multicast frames** and **Sent multicast frames**.

On the user side, the device supports the following query modes for the GPON board:

 Copyright © Huawei Technologies Co., Ltd.

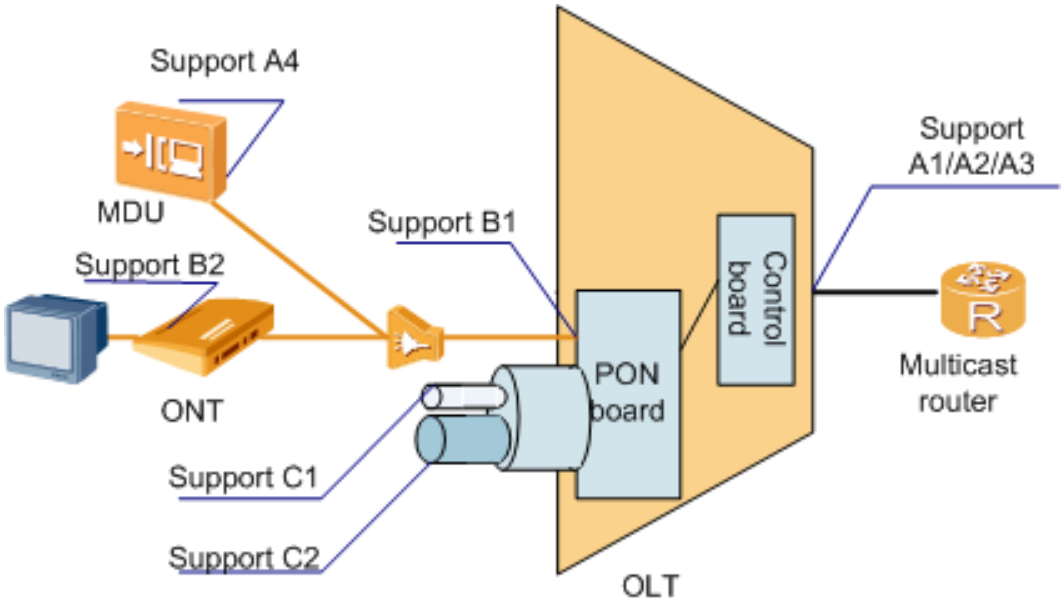
- B1: querying the number of sent/received multicast frames and number of sent bytes on a PON port. Run the **display statistics port ethernet** command to query the number. The command output contains **Received multicast frames** and **Sent multicast frames**.
- B3: querying the number of sent/received multicast frames on an Ethernet port of a specified ONT. Run the **display statistics ont-eth** command to query the number. The command output contains **Received multicast frames** and **Sent multicast frames**.

The following two methods can also be used to query the user-side multicast traffic statistics:

- C1: For the multicast traffic copied to service ports (for P2P boards, and GPON boards that use multi-copy duplication), the traffic statistics (sent/received bytes) of a specified service port can be queried to obtain the multicast traffic statistics. This method is not suitable when a service port carries other services besides the multicast service.
- C2: For P2P boards, or GPON boards that use single-copy duplication, the number of sent packets of queues on a specified port can be queried to obtain the multicast traffic statistics. This method is not suitable when queues on a port carry other services besides the multicast service.

Figure 13-89 shows the points where multicast traffic can be queried.

Figure 13-89 Points of multicast traffic query



Program Ordering Behavior Analysis

Compared with traditional TV services, in the case of IP multicast service, users' order behavior can be measured and analyzed at a finer grain, such as statistics measurement of hottest programs, analysis of user interest, and peak hours of program ordering. For such purposes, the device needs to precisely record the order behavior of each user in the form of logs and output the content of the logs

 Copyright © Huawei Technologies Co., Ltd.

through an open interface. According to different output modes, the device supports two log transfer modes: by CDR or by syslog (RFC 3164). The formats of the two modes are the same. For details, see "**Charging Mode**." The following table lists the pros and cons of the two modes.

Table 13-53 Pros and cons of the two log transfer modes

Log Mode	Pro	Con
CDR	Reliable transfer. TFTP, FTP, or SFTP can be selected as the transfer protocol. **NOTE** FTP and TFTP have security risks due to protocol restrictions. SFTP is recommended.	Logs are reported to the file server only when specified reporting conditions are met (the reporting interval expires or the number of logs reaches the reporting threshold).
syslog	Timely report. Once a log is generated, it is uploaded to the syslog server.	Unreliable transfer. Syslog adopts the UDP protocol.

13.8 Multicast Standards and Protocols Compliance

Table 13-54 Reference standards and protocols

Standard NO.	Standard Description	Application Scope
TR101	Technical Report DSL Forum, TR-101 Migration to Ethernet-Based DSL Aggregation, April 2006	IPv4 and IPv6 Multicast
TR156	Technical Report Broadband Forum, TR-156 Using GPON Access in the context of TR-101, December 2008	IPv4 and IPv6 Multicast
RFC 1112	Deering, S., "Host Extensions for IP Multicasting",STD 5, RFC 1112, August 1989	IPv4 and IPv6 Multicast
RFC-2236	Fenner, W., "Internet Group Management Protocol, Version 2", RFC 2236, November 1997	IPv4 and IPv6 Multicast
RFC 3376	B. Cain., "Internet Group Management Protocol, Version 3 ", RFC 3376,October 2002	IPv4 and IPv6 Multicast
RFC 3569	S. Bhattacharyya, "An Overview of Source-Specific Multicast (SSM)", RFC 3569, July 2003	IPv4 and IPv6 Multicast

 Copyright © Huawei Technologies Co., Ltd.

Standard NO.	Standard Description	Application Scope
RFC 4601	B. Fenner, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", RFC4601, August 2006	IPv4 and IPv6 Multicast
RFC 4604	H. Holbrook, "Using IGMPv3 and MLDv2 for Source-Specific Multicast", RFC 4604, August 2006	IPv4 and IPv6 Multicast
RFC 4605	B. Fenner, "IGMP/MLD Proxying", RFC 4605, August 2006	IPv4 and IPv6 Multicast
RFC 4607	H. Holbrook , "Source-Specific Multicast for IP", RFC 4607, August 2006	IPv4 and IPv6 Multicast
RFC 4541	M. Christensen, "Considerations for IGMP and MLD Snooping Switches", RFC 4541, May 2006	IPv4 and IPv6 Multicast
RFC 2710	S. Deering,"Multicast Listener Discovery (MLD) for IPv6",RFC 2710, October 1999	IPv6 Multicast
RFC 3810	R. Vida, Ed.,"Multicast Listener Discovery Version 2 (MLDv2) for IPv6", RFC 3810, June 2004	IPv6 Multicast

13.9 Multicast Feature Updates

Product Version	Change Description
V100R022C00	Added **13.6.6 Configuring the Layer 2 Reverse Multicast Service** and **13.6.7 Configuring the Layer 3 Reverse Multicast Service**.
V100R020C10	• Added **13.4.2.2 PIM-SM** • **13.7.5 eMDI Video Quality Monitoring for the Unicast Service** supported configuring the mode of reporting the monitoring data to the server.
V100R020C00	Added **13.7.5 eMDI Video Quality Monitoring for the Unicast Service**
V100R019C20	The IPv6 multicast preview function is added.

 Copyright © Huawei Technologies Co., Ltd.

14 Clock and Time

14.1 Clock and Time Feature Updates

14.2 Clock and Time Overview

14.3 Physical Layer Clock/Time Synchronization
This section describes the definition and principle of the physical clock and time system of the Access devices, and describes the specific applications of physical clock and time synchronization.

14.4 1588v2 (based on IEEE)
The IEEE organization proposed the IEEE 1588v2 precision time synchronization protocol, which supports system-wide synchronization accuracy within the sub-microsecond range. In contrast to the GPS solution, the IEEE 1588v2 solution achieves the same time precision but has advantages in terms of cost, maintenance, and security. It has become the most popular time synchronization protocol in the industry.

14.5 1588v2 (Based on G.8275.1)
G.8275.1 defines precision time protocol telecom profile for phase/time synchronization with full timing support from the network based on the IEEE 1588v2 precision time synchronization protocol.

14.6 1588 ACR
1588 Adaptive Clock Recover (ACR) enables a master device that supports IEEE 1588v2 to encapsulate the local system clock information into 1588v2 packets and transparently send them to peer slave devices through a third-party network. The slave devices obtain the timestamp from the 1588v2 packets and restore the clock to implement frequency synchronization between devices at both ends of a packet switched network (PSN).

14.7 NTP
The Network Time Protocol (NTP) is used to synchronize the time between the distributed time server and the client.

 Copyright © Huawei Technologies Co., Ltd.

14.1 Clock and Time Feature Updates

Version	Change Description
V100R019C10	Unified version release is implemented since V100R019C10.

14.2 Clock and Time Overview

14.2.1 Network Synchronization Requirements

Definition

IP-nization is the trend of future network and service development. Difficulties, however, currently exist in the transition from the SDH-based traditional network to the IP-based Ethernet bearer network. One key technology involved is how to bear traditional TDM service on the new network. Traditional TDM service has two major applications: voice service and clock synchronization service.

In a traditional communications network architecture, the TDM service of the fixed network is mainly voice service. Cumulative inconsistency between the clocks at both ends of the bearer network over a long time causes bit slip. The ITU-T Recommendation G.823 defines the requirements on and the test standards of the TDM service of the fixed network. The definition is called the G.823 traffic interface standard. Apart from the bearer network, a traditional communications network usually contains an independent clock-issuing network, which adopts PDH/SDH for issuing clock signals. As specified by the ITU-T, the clock must meet the G.823 Synchronization interface requirements.

In a communications network, the wireless application has the most rigorous requirements on the clock frequency. The frequencies of different BTSs must be synchronized within specified precision. Otherwise, re-synchronization occurs during the BTS switching. Current wireless technologies are in different systems. Different systems have different requirements on the clock bearing. European systems, of which the GSM/WCDMA is a representative, adopt the asynchronous base station technologies. In this case, only frequency synchronization is required, at a precision of 0.05 ppm (or 50 ppb). The clock needs to be provided by the bearer network. The traditional solution is to provide the clock through PDH/SDH. After the IP-nization, the clock needs to be provided by the IP network. The synchronous BTS technologies, of which the CDMA/CDMA2000/TD-SCDMA are representatives, require phase synchronization of the clock (also called time synchronization). **Table 14-1** lists the detailed requirements on clocks.

NOTE

Clock synchronization is frequency synchronization.

Time synchronization is phase synchronization, which requires both phase synchronization and frequency synchronization at the same time.

 Copyright © Huawei Technologies Co., Ltd.

Table 14-1 Requirements of different standards on the clock and time

Wireless System	Frequency Precision	Phase Synchronization Precision
GSM	0.05 ppm	N/A
WCDMA	0.05 ppm	N/A
TD-SCDMA	0.05 ppm	±1.5 us
CDMA2000	0.05 ppm	±3 us
WiMax FDD	0.05 ppm	N/A
WiMax TDD	0.05 ppm	±0.5 us
LTE TDD	0.05 ppm	±1.5 us

Purpose

The purpose is to ensure the clock and time synchronization between communications devices and communications networks.

14.2.2 Synchronization Overview

Synchronization is classified into clock synchronization and time synchronization. Clock synchronization is also called frequency synchronization, and time synchronization is also called phase synchronization.

The synchronization methods are used depending on system types. For SDH networks, frequency synchronization is used, requiring that NEs on the network synchronize their frequencies to ensure normal transmission of SDH services. Wireless stations form a time synchronization system, requiring time synchronization between neighboring stations to ensure switching among stations.

Clock Synchronization

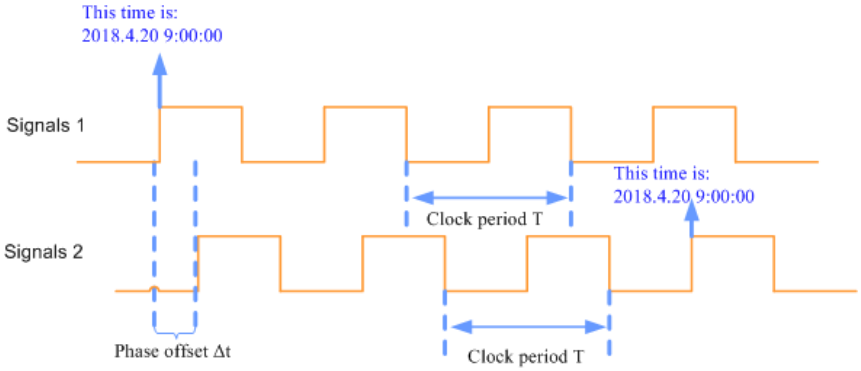
Clock synchronization, also called frequency synchronization, means that signals maintain a specific relationship in frequency or phase. Significant instants of a clock signal occur at the same rate to ensure that equipment in the entire communication network runs at the same rate. In other words, signals maintain a constant phase offset.

Figure 14-1 shows two clocks that synchronize their frequencies to each other. The relationship between the two clocks is summarized as follows:

- The two clocks have the same clock frequency (their clock periods Ts are the same).
- The clock pulses of the two clocks may have different phases (the phase difference is not 0).

 Copyright © Huawei Technologies Co., Ltd.

Figure 14-1 Frequency synchronization



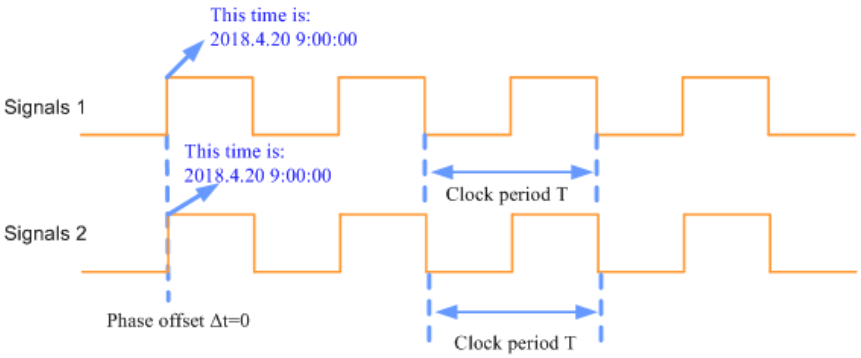
Time Synchronization

Time synchronization, also called phase synchronization, means that signals have the same frequency and phase. That is, there is no phase offset between signals.

Figure 14-2 shows two clocks that synchronize their phases to each other. The relationship between the two clocks is summarized as follows:

- The two clocks have the same clock frequency (their clock periods Ts are the same).
- Clock pulses of the two clocks have the same phase (the phase difference is 0).

Figure 14-2 Phase synchronization



14.2.3 Clock Synchronization

Table 14-2 Comparison between clock synchronization solutions

Clock Source	Description	Remarks
2 MHz or 2 Mbit/s BITS clock	Traces a BITS input clock source and supports 2.048 Mbit/s bitstream (HDB3) or 2.048 MHz clock signal input.	This solution features high deployment costs because each CO requires a BITS device.

 Copyright © Huawei Technologies Co., Ltd.

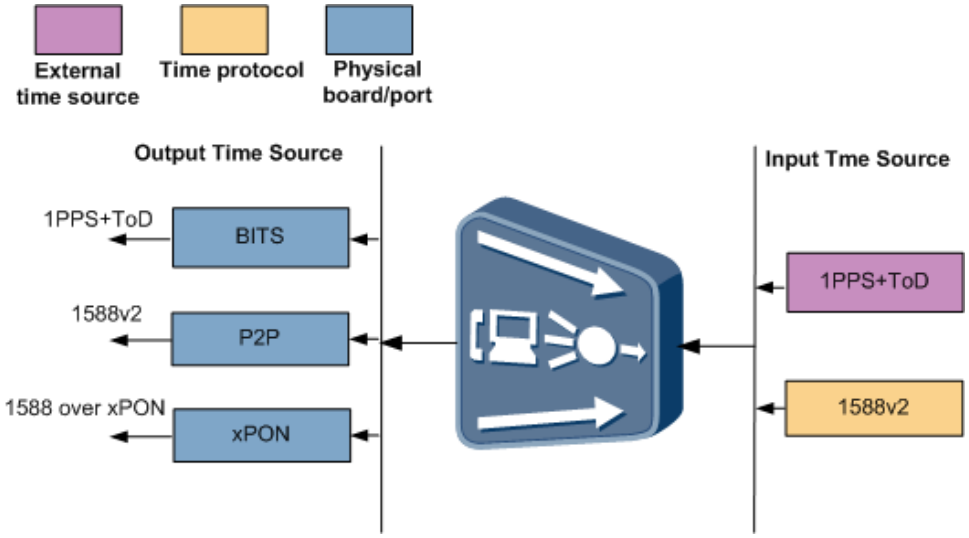
Clock Source	Description	Remarks
Synchronous Ethernet	Synchronous Ethernet uses Ethernet bitstreams to recover clock signals for Ethernet clock synchronization. The implementation mode is similar to that on an SDH or PDH network. The device uses high-precision clock signals as the transit reference in the transmit direction. It recovers and extracts the clock signals at the receive end. The physical layer transmits and receives the clock signals and is compatible with traditional networks. The OLT traces the clock source recovered from GE or 10GE lines. It can also use the clock recovered from line bitstreams as the system clock source.	This solution requires bearer networks to support synchronous Ethernet.
1588v2	The IEEE organization proposed the IEEE1588 V2 precision time synchronization protocol, which supports system-wide synchronization accuracy within the sub-microsecond range. In contrast to the GPS solution, the IEEE1588 V2 solution achieves the same time precision but has advantages in terms of cost, maintenance, and security. It has become the most popular time synchronization protocol in the industry.	This solution requires bearer networks to support IEEE1588 V2 at each hop and applies in mobile bearing scenarios that require time synchronization, such as CDMA2000, TD-SCDMA, and LTE.

 Copyright © Huawei Technologies Co., Ltd.

Clock Source	Description	Remarks
1588 adaptive clock recovery (ACR)	In 1588 ACR, a master device supporting IEEE1588 V2 encapsulates the local system clock data into 1588 V2 packets. The PSN network transparently transmits the 1588 V2 packets to a peer slave device. The slave device obtains timestamps from the received 1588 V2 packets and recovers the clock data to synchronize its frequency with the master device. In the 1588 ACR solution, the PSN network does not need to support clock synchronization or the clock of the PSN network can be a third-party clock, and clock packets are transparently transmitted as service packets.	• Provides E2E clock synchronization and is simple in deployment, meeting clock frequency requirements of MSAN and wireless networks. • Promotes packet delay variation (PDV) requirements on intermediate networks.
PON line clock	Traces the clock source recovered from PON lines and can also use the clock recovered from line bitstreams as the system clock source.	None

14.2.4 Time Synchronization

Figure 14-3 Time synchronization solutions



 Copyright © Huawei Technologies Co., Ltd.

Table 14-3 Comparison between time synchronization solutions

Time Source	Description	Remarks	Time Input	Time Output
1 PPS +TOD	Traces the input time source and supports 1 PPS+TOD time synchronization.	In some mobile bearing scenarios, some intermediate devices do not support time transfer. Therefore, the 1 PPS+TOD time must be sent to the OLT/MDU and the device uses the 1 PPS+TOD time as the system time source. This solution features high deployment costs because each CO requires BITS and GPS devices.	Yes	Yes
1588v2 (based on IEEE)	The IEEE organization proposed the IEEE1588 V2 precision time synchronization protocol, which supports system-wide synchronization accuracy within the sub-microsecond range. In contrast to the GPS solution, the IEEE1588 V2 solution achieves the same time precision but has advantages in terms of cost, maintenance, and security. It has become the most popular time synchronization protocol in the industry.	This solution requires bearer networks to support 1588v2 at each hop and applies in mobile bearing scenarios that require time synchronization, such as CDMA2000, TD-SCDMA, and LTE.	Yes	• 1588v2 over PON • 1588v2 over GE

 Copyright © Huawei Technologies Co., Ltd.

Time Source	Description	Remarks	Time Input	Time Output
1588v2 (based on G.8275.1)	G.8275.1 defines precision time protocol telecom profile for phase/time synchronization with full timing support from the network based on the IEEE 1588v2 precision time synchronization protocol.	This solution requires bearer networks to support G.8275.1 hop by hop and applies in mobile bearing scenarios that require time synchronization, such as CDMA2000, TD-SCDMA, and LTE.	Y	• 1588v2 over PON • 1588v2 over GE/10GE

14.3 Physical Layer Clock/Time Synchronization

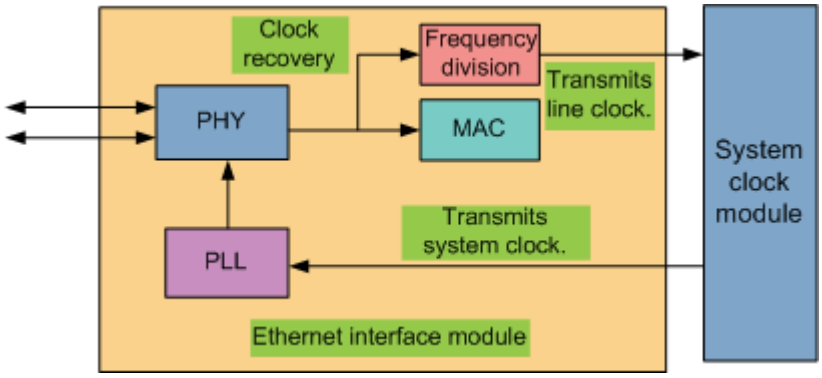
This section describes the definition and principle of the physical clock and time system of the Access devices, and describes the specific applications of physical clock and time synchronization.

14.3.1 Physical Layer Clock/Time Synchronization Principles

14.3.1.1 Synchronous Ethernet

Synchronous Ethernet, similar to traditional SDH synchronization, is a synchronization technology that uses physical-layer bitstreams to carry frequency information and recovers frequency information also from physical-layer bitstreams. Synchronous Ethernet enables the device to extract clock signals from serial bitstreams on an Ethernet line, select a better clock source, and finally send the clock signals (line clock) to a system clock phase-locked loop (PLL). The PLL then traces the clock signals and generates a system clock and uses the system clock as the transmit clock at the Ethernet physical layer to transmit clock data to lower-layer devices.

Figure 14-4 Synchronous Ethernet clock signal processing



 Copyright © Huawei Technologies Co., Ltd.

- Ethernet interface receive clock:
  - The PHY chip in the Ethernet interface module recovers a clock from Ethernet line bitstreams, divides the frequency of the clock, and sends the clock to the system clock module.
  - The system clock module selects the clock source with the highest precision based on clock source priorities and sends the clock source to the system PLL as the system reference clock. Then, the PLL traces the reference clock and outputs a high-precision clock for each interface.
- Ethernet interface transmit clock:
  - The system clock module sends a high-precision clock to Ethernet interfaces on NEs.
  - The PLL of the Ethernet interface module traces the high-precision clock sent from the system clock module, generates a reference clock for the Ethernet link bitstreams, and sends the reference clock out using Ethernet link bitstreams.

Performance and Deployment Limitations

A synchronous Ethernet clock is simple to implement. The mechanism for transferring the clock in synchronous Ethernet is simple to implement and the recovered clock is reliable in compliance with ITU-T G.823 with respect to the synchronization interface specifications.

Similar to an SDH network, a synchronous Ethernet network has limitations on deployment. For the synchronous Ethernet network, the clock is transferred over links. All nodes on the clock trail must have the synchronous Ethernet feature to achieve clock synchronization (only frequency synchronization) of the entire network.

Synchronous Ethernet can work with IEEE 1588v2 clocks for time synchronization.

14.3.1.2 External Time Source

A time source is a signal source containing reference timing information. Each NE synchronizes its local clock phase to the reference timing using its phase-locked loop (PLL). In this manner, time synchronization is implemented on the entire network.

A time source can be from an external time source or a line time source.

In some mobile carrying scenarios, 1PPS+TOD time signals must be injected to an OLT or an MDU because intermediate devices do not support transmission of these signals. Then the signals are used as the system time source.

The 1PPS+TOD protocol receives signals using a solution similar to time signal reception using a GPS satellite. This protocol transmits two types of signals: 1PPS and TOD signals. Frame headers and time information are transmitted over dual lines.

1PPS+TOD time signals consist of 1PPS signals and TOD time information.

- 1PPS

  1PPS is short for one pulse per second. 1PPS signals are used for time scaling and work at the RS-422 levels. The pulse frequency of 1PPS is 1 Hz. That is,

 Copyright © Huawei Technologies Co., Ltd.

one pulse is transmitted per second. The 1PPS signal pulse width ranges from 20 ms to 200 ms. On the rising edge of the pulse, UTC time signals are aligned.

- TOD

  TOD is short for time of day. TOD messages provide time in ASCII format. TOD signals also work at the RS-422 or RS-232 levels and provide a baud rate of 9600 bit/s. A TOD message contains information such as current date/time, time standard ID, 1PPS status flag, date/time adjusted based on UTC leap seconds, leap second adjustment directive, and GPS time.

TOD can be in UBX or NMEA format.

Table 14-4 TOD Format

TOD Format	Description
can be in UBX	a private format of the Ublox company. The UBX format is in binary code and identified by a length domain because it does not contain a start or end character.
NMEA	a set of communication protocol formulated by the National Marine Electronics Association to standardize various global positioning system (GPS) navigators. The NMEA protocol is in ASCII format and identified by start and end characters. A comma (,) is used to separate domains.

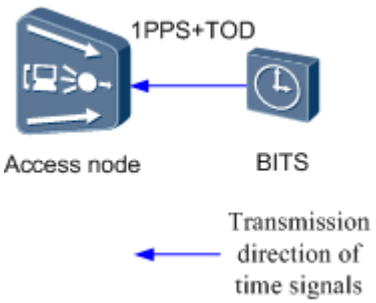
NOTE

The OLT supports only the UBX format.

As shown in **Figure 14-5**, the OLT or MDU can function as an access node to receive 1PPS+TOD clock signals.

- When the OLT functions as the access node, it receives 1PPS+TOD clock signals using the BITS/TOD port as the clock source.
- When the MDU functions as the access node, it receives 1PPS+TOD clock signals using the clock/time output port BITS/TOD port as the clock source.

Figure 14-5 External time source synchronization



 Copyright © Huawei Technologies Co., Ltd.

14.3.1.3 Clock Mode

This topic describes the three clock modes supported: Locked mode, Holdover mode and Free-run mode.

Locked Mode

In this mode, the system clock source is synchronized with the input clock source. The phases of the system clock source and the input clock source are in a constant relationship. The device traces the BITS clock source, upstream Ethernet synchronous clock, line clock source, Auto-adaptation recovery clock, or IEEE1588 V2 message recovered clock source. The ideal clock source can meet the 50 ppb requirements of mobile backhaul.

Holdover Mode

The device records the clock data of the trace mode. If the clock source is lost, the system builds a system clock by using the recorded clock data, and maintains the clock properties as consistent as possible with the clock properties in the trace mode. As such, the system enters and remains in the holdover mode. The precision of the holdover mode meets the G.813 or G.8262 requirements.

NOTE

The system supports the holdover mode only when configured with the stratum-3 clock daughter board.

Free-Run Mode

In this mode, the device works based on the inherent frequency of the internal crystal oscillator inside the device.

14.3.1.4 Clock Source Selection

The device uses a clock source selection algorithm to select a clock source as its reference clock. It supports the clock source selection algorithm based on clock source priorities or synchronization status messages (SSMs).

Clock Source Priority-based Clock Source Selection

If uses the clock source priority-based clock source selection algorithm to select a clock source, users need to configure the external clock sources with different priorities. The OLT/MDU then automatically selects the clock source that has the highest priority and is running properly as its reference clock. When the traced clock source becomes faulty, the device automatically selects the clock source that has the second highest priority and is running properly as its reference clock.

SSM-based Clock Source Selection

SSM messages are used for transporting clock signal quality levels over clock links. By reading data carried in the SSM messages, the node clocks on SDH and synchronization networks obtain the information of the upper-layer clocks, perform operations based on instructions carried in the SSM messages, such as

 Copyright © Huawei Technologies Co., Ltd.

tracing a clock, switching a clock source, or entering the holdover state, and send the clock synchronization information of the node clocks to lower-layer clocks.

The following table lists ITU-T G.781-compliant SSM quality levels.

Quality Level	SSM Code	Code Value	Description
QL-PRC	0010	0x02	Indicates that the timing quality generated by a primary reference clock defined in Recommendation G.811 is transported.
QL-SSU-A	0100	0x04	Indicates that the timing quality generated by a type I or V (transit exchange) slave clock defined in Recommendation G.812 is transported.
QL-SSU-B	1000	0x08	Indicates that the timing quality generated by a type VI (local exchange) slave clock defined in Recommendation G.812 is transported.
QL-SEC	1011	0x0b	Indicates that the timing quality generated by an SDH equipment clock (SEC) defined in Recommendation G.813 is transported.
QL-DNU	1111	0x0f	Indicates that the clock source is not used for synchronization.

If the SSM-based clock source selection algorithm is used, the device extracts the quality level from clock source input signals. If the clock source does not support SSM information extraction, users must manually configure the quality level. The device automatically selects the clock source that has the highest quality level and is running properly as its reference clock. Ensure that the quality level is higher than or equal to the lowest SSM level of the clock source.

- If there are two external clock sources with the same highest quality level, the device selects the one that has a higher clock source priority.
- If the clock source traced by the OLT/MDU becomes faulty, the OLT/MDU automatically selects the clock source that has the second highest quality level and is running properly as its reference clock.

In SSM-based clock source selection mode, SSM quality levels of clock sources are used for selecting a clock source. The SSM quality levels of clock sources can be obtained in either of the following modes:

- Users configure the SSM quality levels of clock sources. In this mode, all clock source ports support SSM quality level configurations. For the clock source ports that support the extraction of SSM quality levels, users can also

 Copyright © Huawei Technologies Co., Ltd.

configure SSM quality levels for the clock source ports as the SSM quality levels of these clock sources.

- The device extracts SSM information, from lines. Not all boards support the extraction of SSM information from lines. For details, see section **Feature Dependencies and Limitations**.

The device uses Ethernet (SSM), E1, , or PON ports to transport clock quality levels to lower-layer devices.

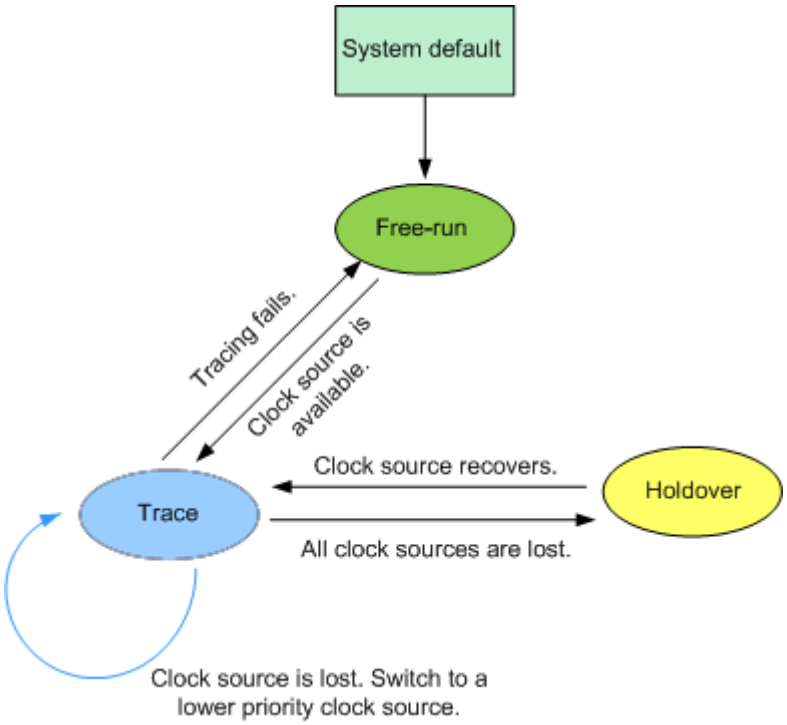
- E1 (2 Mbit/s) ports use four idle bits in the TS0 timeslot to transport quality levels.
- Ethernet clock synchronization ports use ESMC packets to transport quality levels.

The device enters holdover if no clock source is available, regardless of whether the clock source selection algorithm is based on clock source priorities or SSM messages. In holdover mode, the device uses the data recorded in tracing mode to control the clock unit so that the clock unit generates clock properties similar to those of the clock in tracing mode.

- If clock recovery is enabled, the device returns to the tracing mode from the holdover mode when a clock source recovers.
- If the clock recovery mode is disabled, the device does not return to the tracing mode no matter whether a clock source recovers in the holdover period.
- If the device restarts during the holdover period, the system enters the free-run mode.

Figure 14-6 shows the clock source status transition.

Figure 14-6 Clock source status transition



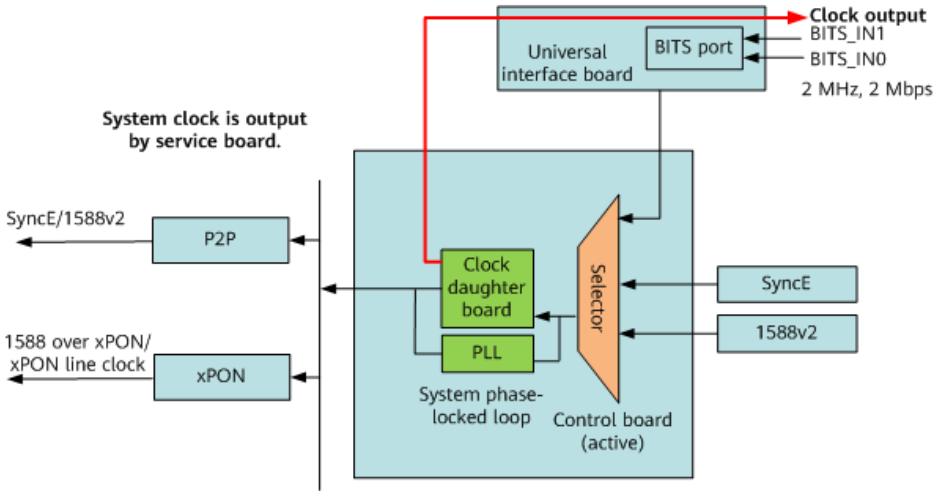
 Copyright © Huawei Technologies Co., Ltd.

14.3.2 Physical Layer Clock/Time Synchronization Usage Scenarios

14.3.2.1 Applications of Clock Output

By specific configuration, the OLT can select a system clock output. The output clock can serve as the clock source for other devices, as shown in **Figure 14-7**.

Figure 14-7 Clock output



When the system clock is selected as the output clock, the clock output through the universal interface board port is the system clock phase-locked by the MA5800.

The following clock sources can serve as the system clock:

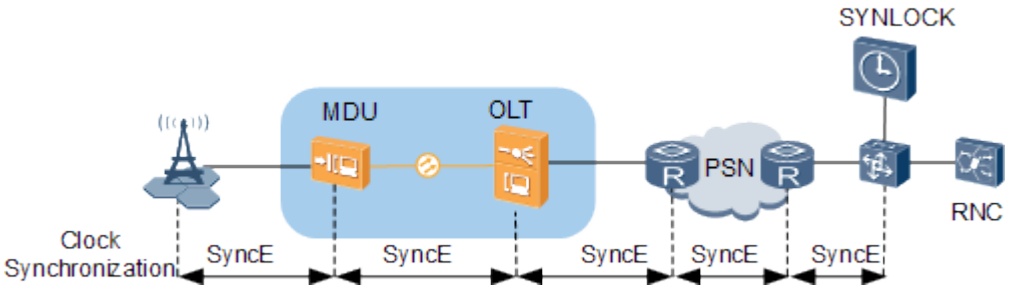
- BITS clock source: including 2 Mb/s and 2 MHz signals.
- Synchronous Ethernet line clock
- IEEE1588 V2 clock source
- Free-running internal clock source

14.3.2.2 Clock Synchronization of the Synchronous Ethernet Service

Traditional Ethernet application does not consider the synchronization requirement. The Ethernet ports adopt the ±100 ppm local oscillator as the transmit clock, and the transmit clocks of the NEs are independent of each other. As such, the clocks are not precise enough. Synchronous Ethernet is a technology that recovers the clock from the bit streams on the Ethernet link and implements synchronization between Ethernet networks. The implementation mode is similar to the synchronization mode on the SDH/PDH network. In the transmit direction, the high-precision system clock is adopted as the transmit clock, which is recovered and obtained at the receive end. The transmission and reception are performed by the physical layer independently, which in terms of function is compatible with traditional Ethernet.

 Copyright © Huawei Technologies Co., Ltd.

Figure 14-8 Clock synchronization for the synchronous Ethernet service



14.3.3 Configuring the Physical Clock

On a digital network comprising the access device and other devices, the primary problem to be solved is clock synchronization for carrying the traditional TDM service. To ensure the system clock synchronization of each device in the digital network, a system clock source must be specified.

Context

IP-based solution is the trend of future network and service development, so is the trend of the bearer network. Difficulties, however, currently exist in the transition from the SDH-based traditional network to the IP-based Ethernet bearer network. One key technology involved is how to carry traditional TDM service on the new network. Traditional TDM service has two major applications: voice service and clock synchronization service. In a traditional communications network architecture, the TDM service of the fixed network is mainly voice service. Cumulative inconsistency between the clocks at both ends of the bearer network over a long time causes frame slip. On a communications network, the wireless application has the most rigorous requirements on the clock frequency. The frequencies of different base stations must be synchronized within a specified precision. Otherwise, re-sync occurs during the base station switching.

To ensure clock synchronization among devices, relevant clock synchronization methods are adopted based on the clock source solution provided by an upper-layer device.

14.3.3.1 Configuring the System Clock Based on the Priority

If a device has multiple clock sources and the precision of the clock sources is provided, you need to configure the priorities of the clock sources. Generally, the higher the precision is, the higher the priority is.

Context

A clock source can be an external BITS clock or a line clock from the upper-layer node. The clock module automatically judges the types of the specified clock sources (BITS, TDM), and sends them according to their priorities to the clock module, serving as clock sources for phase lock.

 Copyright © Huawei Technologies Co., Ltd.

Procedure

Step 1 Run the **clock source sourceid** command to configure the system clock source.

Specify the clock signals extracted from a certain port as the system clock source.

- The system supports 10 clock sources in total.
- Only the external clock sources on the physical entities are added by running this command and the external clock sources are numbered. To enable the relevant external clock source, you need to run the **clock priority system** command to determine whether the relevant clock source is available.
- The system clock cannot serve as the system clock source.

Step 2 Run the **clock priority system** command to configure the priority of the system clock source.

- The system supports 10 clock source priorities. The highest priority is p0 and the lowest priority is p9.
- When the clock source is selected based on the priority, the system does not check the quality of the clock source. Therefore, you must configure the clock source of high quality with a high priority.
- After the priority of the clock source is configured, the system selects the clock source with the highest priority and in the normal state as the system clock source.
- When the clock source with the highest priority is faulty, the system automatically switches to the clock source with the second highest priority.
- When the clock source with the highest priority recovers, the system switches back to this clock source.

----End

Example

Assume the following configurations: The device obtains three clock sources from a BITS port on the universal interface board and ports 0/2/0 and 0/2/1 on the upstream interface board as clock source 0, clock source 1, and clock source 2 of the system. Configure clock source 0 with the highest priority and configure clock source 2 with the lowest priority. To perform the preceding configurations, do as follows:

```
huawei(config)#clock source 0 0/0/0 2MHz 120ohm
huawei(config)#clock source 1 0/2/0
huawei(config)#clock source 2 0/2/1
huawei(config)#clock priority system 0/1/2
```

14.3.3.2 Configuring the System Clock Based on the SSM Clock Source Selection Mode

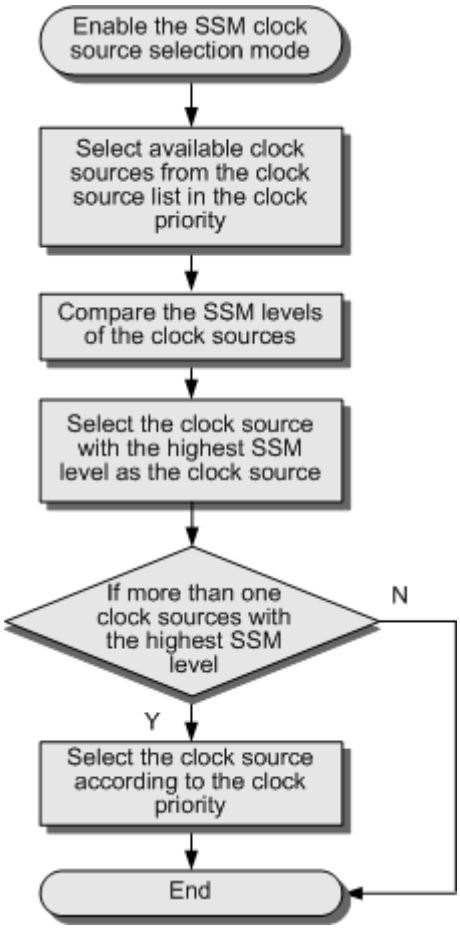
If the clock transmitted from an upper-layer device contains a synchronization status message (SSM) and all clock sources are selected based on the SSM, you need to configure the system clock based on the SSM clock source selection mode.

 Copyright © Huawei Technologies Co., Ltd.

Context

By default, the SSM clock source selection mode is disabled. That is, the system selects the clock source based on the priority. The system enables the SSM clock source selection mode only after the system determines that the clock source contains an SSM and the entire system are based on the SSM clock source selection mode.

For the detailed SSM clock source selection process, see the following flowchart.



Procedure

Step 1 Enable the SSM clock source selection mode.

Run the **clock ql-mode enable** command to enable the SSM clock source selection mode.

Step 2 Run the **clock source** command to configure the system clock source.

Step 3 Run the **clock priority** command to configure the clock source range based on the SSM clock source selection mode and the priority sequence of the corresponding clock sources if the SSM quality levels are the same.

Step 4 (Optional) Run the **clock ql sourceid clock-ql** command to configure the SSM quality level for the clock source. If the SSM clock source selection mode is enabled, but a clock source does not support the output of an SSM, you need to

 Copyright © Huawei Technologies Co., Ltd.

manually configure the SSM quality level for the clock source. After the SSM quality level is configured, the device no longer matches the received SSM.

NOTE

- When the SSM clock source selection mode of the system is disabled, the SSM quality level of the clock source cannot be set.
- When the system selects the clock source based on the SSM quality level, the system selects the clock sources based on the priority and then compares the SSM of the clock sources. Finally, the system selects the clock source with the highest SSM quality level as the system clock source. If there are multiple clock sources with the same SSM qualify level, the system selects the clock source based on the priority.

Step 5 (Optional) Run the **clock ql input lower-limit** command to configure the lowest synchronization status message (SSM) quality level threshold of the clock source. When the SSM quality level of the clock source of the upper-layer device is greater than or equal to that of the device, the clock of the upper-layer device is traced. Otherwise, the clock of the device is in the holdover state and switched to the free-run state after 24 hours elapses.

Step 6 (Optional) Run the **clock ql output** command to configure whether the specified port sends the SSM quality level.

----End

Example

Example 1:

Configure the SSM clock source selection mode as the system clock source selection mode. Configured the three clock sources obtained from upstream ports **0/9/0** and **0/9/1** on the control board and port 0/2/0 on the upstream board as clock source 0, clock source 1, and clock source 2 with SSM respectively. Configure clock source 0 with the highest priority and clock source 2 with the lowest priority and set the lower threshold to QL-SSU-B for received clock sources. To perform the preceding configurations, do as follows:

```
huawei(config)#clock ql-mode enable
huawei(config)#clock source 0 0/9/0
huawei(config)#clock source 1 0/9/1
huawei(config)#clock source 2 0/2/0
huawei(config)#clock priority system 0/1/2
huawei(config)#clock ql input lower-limit QL-SSU-B
```

Example 2:

Configure the SSM clock source selection mode as the system clock source selection mode. Configure the clock sources obtained from upstream ports **0/9/0** and **0/9/1** on the control board as clock source 0 and clock source 1 with SSM respectively; and the clock source obtained from port 0/0/0 (that does not support SSM) on the universal interface board as clock source 2. Configure clock source 0 with the highest priority and clock source 2 with the lowest priority.

Manually set the SSM quality level to QL-SSU-A for clock source 2 (port 0/0/0) that does not support SSM and set the lowest threshold for the received clock source to QL-SSU-B.

Send SSM from port 0/3/1. To perform the preceding configurations, do as follows:

```
huawei(config)#clock ql-mode enable
huawei(config)#clock source 0 0/9/0
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#clock source 1 0/9/1
huawei(config)#clock source 2 0/0/0 2MHz 120ohm
huawei(config)#clock ql 2 QL-SSU-A
huawei(config)#clock priority system 0/1/2
huawei(config)#clock ql input lower-limit QL-SSU-B
huawei(config)#clock ql output 0/3/1 enable
```

14.3.3.3 Configuring External Clock

The device can select a system clock output or export the line clock as the clock source of another device.

Context

The external clock output supports the selection of the following two benchmark clocks.

- Select the system clock as the output benchmark clock.
- Select the line clock as the output benchmark clock.

Procedure

Step 1 Run the **clock external bits-type** command to set the signal type of the BITS port.

NOTE

The system defaults to 2 Mbit/s.

Step 2 Configuring the clock source of the external clock

1. Run the **clock source sourceid** command to configure the clock sources of the external clock.

Step 3 Run the **clock priority systemian** command to configure the priority of the system clock source.

- The system supports 10 clock source priorities. The highest priority is p0 and the lowest priority is p9.
- When the clock source is selected based on the priority, the system does not check the quality of the clock source. Therefore, you must configure the clock source of high quality with a high priority.
- After the priority of the clock source is configured, the system selects the clock source with the highest priority and in the normal state as the system clock source.
- When the clock source with the highest priority is faulty, the system automatically switches to the clock source with the second highest priority.
- When the clock source with the highest priority recovers, the system switches back to this clock source.

----End

Example

Assume the following configurations: Configure the SyncE clock sources obtained from port 0 on the control board in slot **0/9** and port 0 on the upstream interface board in slot 0/2 as clock source 1 and clock source 2 of the system respectively. Configure clock source 1 with a higher priority than clock source 2, and configure

 Copyright © Huawei Technologies Co., Ltd.

the system clock as the output clock. To perform the preceding configurations, do as follows:

```
huawei(config)#clock source 1 0/9/0
huawei(config)#clock source 2 0/2/0
huawei(config)#clock priority system 1/2
```

14.3.4 Physical Layer Clock/Time Synchronization Standards and Protocols Compliance

The following lists the reference standards and protocols of this feature:

- ITU-T G.823 (the control of jitter and wander within digital networks which are based on the 2048 kbit/s hierarchy)

  The ITU-T G.823 describes the jitter and wander of the PDH interface in the 2048 kbit/s synchronous system. Here, the E1 port is mainly referred to, which meets the G.824 requirements for the 1544 kbit/s system. G.823 defines the requirements on the traffic interface and the synchronization interface. The traffic interface performance is the basic requirement for service transmission. To transmit synchronous clock signals, the E1 port must meet the requirements of the synchronization interface. The synchronization interface has stricter requirements on jitter and wander than the traffic interface does.
  - The G.823 defines the following items for the traffic interface and the synchronization interface:
    - Jitter output
    - Output wander
    - Input jitter and wander tolerance
- ITU-T G.825, The control of jitter and wander within digital networks which are based on the synchronous digital hierarchy (SDH)

  The ITU-T G.825 defines the jitter and wander characteristics of SDH devices.
  - The ITU-T G.825 defines the following items for the STM port:
    - Jitter output
    - Output wander
    - Input wander tolerance
    - Input jitter tolerance
    - Jitter and wander generation
    - Jitter and wander transfer
- ITU-T G.8261, Timing and Synchronization Aspects in Packet Networks

  The ITU-T G.8261 defines the wander budget of CES and synchronous Ethernet for packet networks. The ITU-T G.8261 requirements are similar to the requirements of the ITU-T G.823 on the TDM network. The CESoP service and synchronous Ethernet clock feature of the MA5800 need to meet the ITU-T G.8261 requirements.
- ITU-T G.8262, Timing characteristics of synchronous Ethernet equipment slave clock (EEC)

 Copyright © Huawei Technologies Co., Ltd.

The ITU-T G.8262 defines the requirements on the synchronous Ethernet clock system, which are equal to the G.813 and G.812 requirements on the TDM network. To meet the ITU-T G.8262 specifications, the MA5800 must be configured with the stratum-3 clock daughter board.

- ITU-T G.8264, Distribution of timing through packet networks

  The ITU-T Recommendation G.8264 defines the SSM protocol and message format for the synchronous Ethernet to ensure clock synchronization between Ethernet devices.

14.4 1588v2 (based on IEEE)

The IEEE organization proposed the IEEE 1588v2 precision time synchronization protocol, which supports system-wide synchronization accuracy within the sub-microsecond range. In contrast to the GPS solution, the IEEE 1588v2 solution achieves the same time precision but has advantages in terms of cost, maintenance, and security. It has become the most popular time synchronization protocol in the industry.

14.4.1 Why Is 1588v2 Required

Background

1. Traditionally, time synchronization chains use the NTP, which can only ensure time precision within milliseconds. This precision is far below the time synchronization precision of microseconds required by wireless base stations.
2. On the base station side, GPS can be used to solve frequency and time synchronization problems. However, GPS has many problems, such as difficult site selection, high fault rate, difficult maintenance, difficult feeder routing, high security risks, and high costs.
3. To meet the high time precision requirement of wireless synchronization base stations and solve the disadvantages of the existing GPS solution, a high precision time synchronization solution for terrestrial transmission is urgently needed.

IEEE 1588v2 high-precision Time Transfer Protocol

Based on the synchronization requirements for different wireless systems and time synchronization condition, IEEE defines IEEE 1588v2 high-precision time protocol (full name: Precision Clock Synchronization Protocol for Networked Measurement and Control Systems and abbreviation: PTP). The protocol supports time synchronization precision in the sub-microsecond range.

In contrast to the GPS solution, IEEE 1588v2 solution achieves the same time precision but has advantages in terms of cost, maintenance, and security. It has become the most popular time transfer protocol in the industry because of the following advantages:

1. Space localization: The IEEE 1588v2 protocol is applicable to local area networks supporting multicast messaging including but not limited to Ethernet.

 Copyright © Huawei Technologies Co., Ltd.

2. Low cost: The IEEE 1588v2 protocol helps minimize the demand for network resources and CPU resources and therefore enables low-cost applications.
3. Network transformation trend: IP networks are transformed to future—oriented convergence networks.

14.4.2 1588v2 Basic Concepts

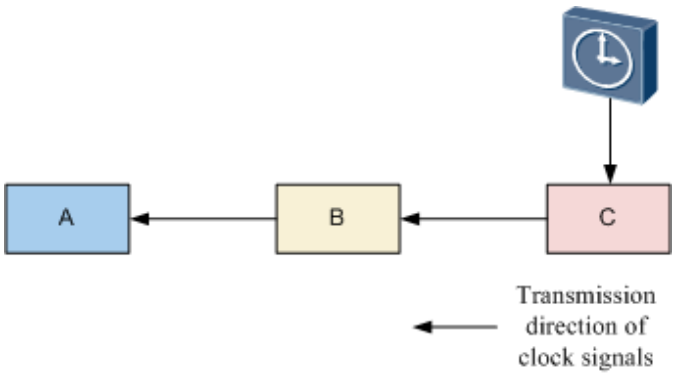
14.4.2.1 Device Model

When a pair of nodes performs time synchronization, the upstream node distributing the reference time signals is the master node and the downstream node receiving the reference time signals is the slave node. As shown in **Figure 14-9**, B is the master node, and A is the slave node.

In the subnet system, grand master is the master clock. Each system has only one grandmaster clock (GMC), and each subnet has only one master clock. Slave clocks are kept synchronized with the master clock.

As shown in **Figure 14-9**, C is the grand master, and nodes A and B must be synchronized with C.

Figure 14-9 Synchronization on devices

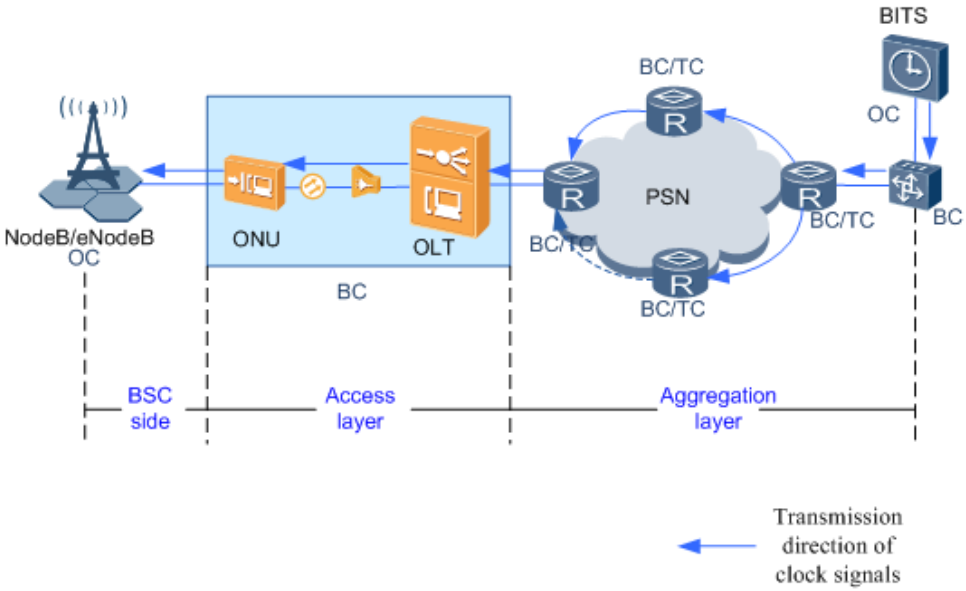


1588v2 Device Model

The IEEE 1588v2 standard defines five network node models: ordinary clock (OC), boundary clock (BC), end-to-end transparent clock (E2E TC), peer-to-peer transparent clock (P2P TC), and management node.

 Copyright © Huawei Technologies Co., Ltd.

Figure 14-10 Device model



OC

An OC has only one 1588v2 clock interface (a clock interface enabled with 1588v2) through which the OC synchronizes with an upstream node or distributes time signals to downstream nodes. An OC has only one physical port. Therefore, in practice, an OC can either be a grandmaster (GM) clock or a slave lock.

As shown in **Figure 14-10**, BITS is generally configured as an OC, functioning as the GM clock in the entire network; a basic station, as a slave device, is also configured as an OC.

BC

A BC has several physical ports for network communication. Each port is similar to the port of an OC, which is used to connect multiple domains. One of the ports is used to synchronize with an upstream node. The other ports can be used to distribute time signals to downstream nodes.

In practice, a BC synchronizes its time with the upstream device and issues the synchronized time to the downstream devices.

As shown in **Figure 14-10**, BC contains the access network devices at the access layer.

E2E TC

An E2E TC provides several physical ports, forwards all PTP messages, and measures and corrects the residence time of PTP event messages traversing the E2E TC.

P2P TC

A P2P TC has multiple ports. In addition to functions of an E2E TC, a P2P TC is also able to compute and correct the link delay between each port and similarly equipped port on another node sharing the link, that is, the link peer.

 Copyright © Huawei Technologies Co., Ltd.

Difference between a TC and BC/OC: BC/OC must be synchronized with the clock on other devices, and TC does not restore the clock from IEEE 1588 packets but processes and forwards IEEE 1588 packets.

As shown in **Figure 14-10**, a TC can function as a device at the aggregation layer.

Management node

A management node has multiple ports and serves as a management interface to PTP management messages. The management node is only used to manage the synchronization nodes and does not provide the synchronization function.

14.4.2.2 Clock Domain and Clock ID

Clock Domain

Logically, a physical network can be divided into multiple clock domains. Each clock domain has a reference time with which all devices in the domain are synchronized. Each clock domain has its own reference time and these times are independent of one another.

A device can join only one clock domain and can synchronize only with the synchronization time of that clock domain.

Clock ID

A clock ID identifies a clock in an IEEE 1588v2 clock subnet. In an IEEE 1588v2 packet, a clock source ID occupies eight bytes. It consists of two parts: The first three bytes are OUI code, and the other five bytes are extended ID.

Organizational Unique Identifier (OUI): an organization identifier uniformly assigned by the IEEE standard.

Extended ID: an identifier uniformly assigned by the organization represented by the OUI to ensure that the clock ID in each IEEE 1588v2 packet is unique.

14.4.2.3 1588v2 Messages

Message Type

As shown in **Table 14-5**, there are two types of IEEE 1588v2 messages: event message and general message.

 Copyright © Huawei Technologies Co., Ltd.

Table 14-5 Message Type

Message Type	Message	Description
Event message	• Sync • Delay_Req • Pdelay_Req • Pdelay_Resp	Event messages are timed messages in that an accurate timestamp is generated both at the device ingress and egress. Timestamps need to be processed in a timely manner at the time of message transmission and reception.
General Message	• Announce • Follow_Up • Delay_Resp • Pdelay_Resp_Follow_Up	General messages are not timed messages and do not require accurate timestamps. Timestamps do not need to be processed in a timely manner at the time of message transmission and reception.

- The Sync, Delay_Req, Follow_Up, and Delay_Resp messages are used to generate and communicate the timing information needed to synchronize OCs and BCs using the delay request-response mechanism.
- The Pdelay_Req, Pdelay_Resp, and Pdelay_Resp_Follow_Up messages are used to measure the link delay between two clock ports implementing the Pdelay mechanism. The link delay is used to correct timing information in Sync and Follow_Up messages in systems composed of P2P TCs. OCs and BCs that implement the Pdelay mechanism can synchronize using the measured link delays and the information in the Sync and Follow_Up messages.
- The Announce message is used to establish the synchronization hierarchy. That is, information related to the best master clock (BMC) algorithm is carried in the Announce messages.

Message Format

An IEEE 1588v2 message must have a header, body, and suffix. The suffix length may be 0.

An IEEE 1588v2 message may be encapsulated in Ethernet or IP format.

IEEE 1588v2 message encapsulated in Ethernet format

Messages encapsulated in Ethernet format are classified into IEEE 1588v2 over Ethernet messages without VLAN tags and IEEE 1588v2 over Ethernet messages with VLAN tags.

 Copyright © Huawei Technologies Co., Ltd.

Figure 14-11 Format of an IEEE 1588v2 over Ethernet message

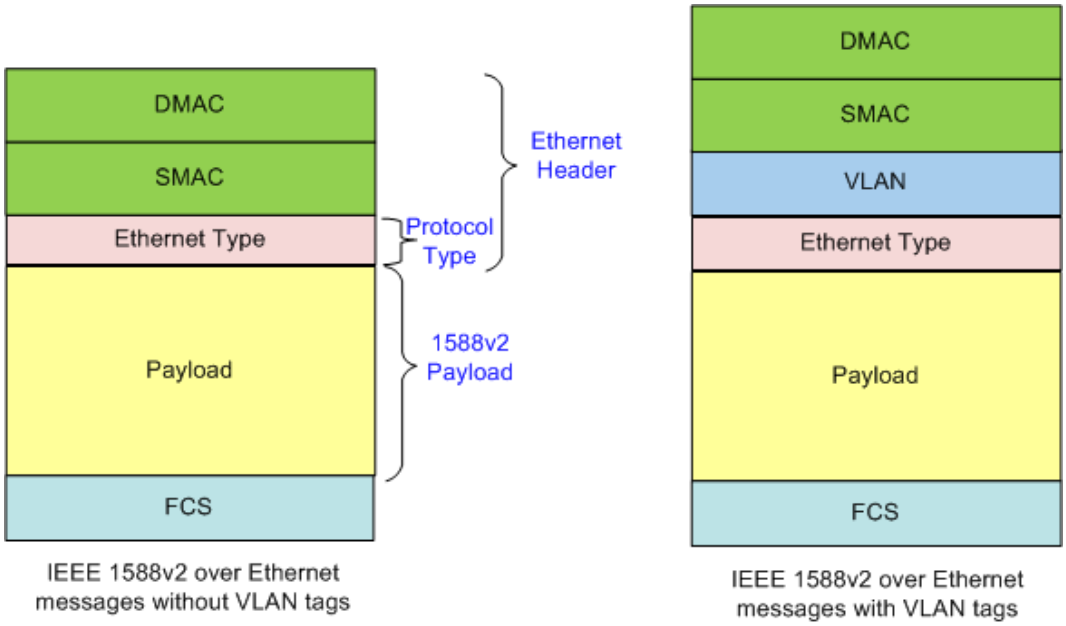


Table 14-6 describes the fields in an IEEE 1588v2 over Ethernet message.

Table 14-6 Fields in an IEEE 1588v2 over Ethernet message

Field Name	Bit Width (Unit: Bit)	Description
DMAC	6 x 8 = 48	Destination MAC address. It is configured by a user. • When the destination MAC address is not configured, the multicast encapsulation mode is used. This field is filled with 01-80-C2-00-00-0E for Pdelay packets and filled with 01-1B-19-00-00-00 for other packets. • When the destination MAC address is configured, the unicast encapsulation mode is used. This field is filled with the configured destination MAC address or with the SMAC of received packets.
SMAC	6 x 8 = 48	Source MAC address. It is configured by a user. It is the MAC address of an NE's control board.
VLAN	4 x 8 = 32	VLAN tag. It is configured by a user. At most one VLAN tag is supported, and TPID can only be 0x8100.
Ethernet Type	2 x 8 = 16	Ethernet type. It is configured by the system. This field has a fixed value of 88F7, which indicates an IEEE 1588v2 over Ethernet message.

 Copyright © Huawei Technologies Co., Ltd.

Field Name	Bit Width (Unit: Bit)	Description
Payload	N = (A value ranging from 44 to 64) x 8	Payload of an IEEE 1588v2 over Ethernet message.
FCS	4 x 8 = 32	CRC32 check value.

IEEE 1588v2 message encapsulated in IP format

Messages in encapsulated in IP format are classified into IEEE 1588v2 over IPv4 messages without VLAN tags and IEEE 1588v2 over IPv4 messages with VLAN tags.

 Copyright © Huawei Technologies Co., Ltd.

Figure 14-12 Format of an IEEE 1588v2 over IPv4 message

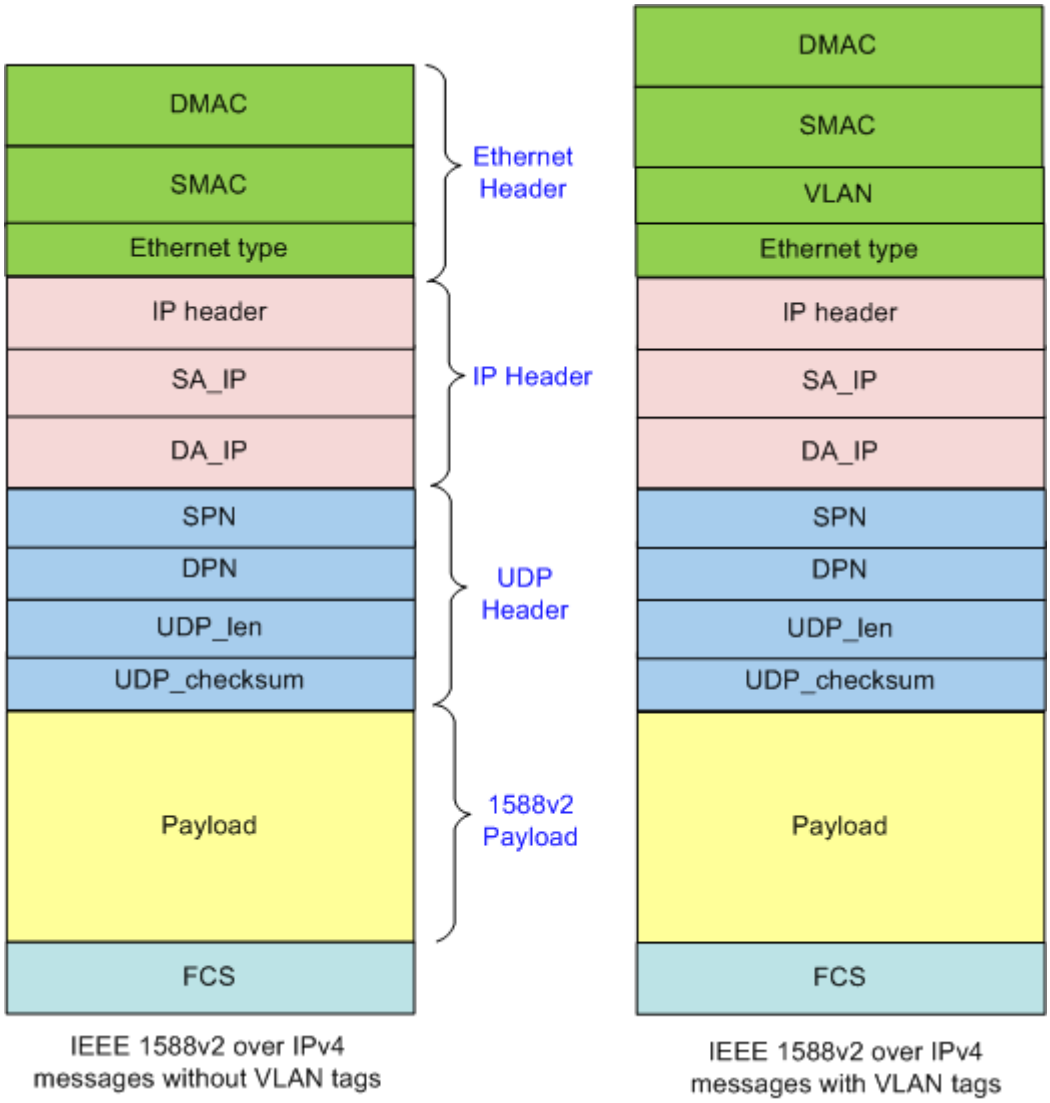


Table 14-7 lists the fields in an IEEE 1588v2 over IPv4 message.

Table 14-7 Fields in an IEEE 1588v2 over IPv4 message

Field Name	Bit Width (Unit: Bit)	Description
DMAC	6 x 8 = 48	-
SMAC	6 x 8 = 48	-
Ethernet Type	2 x 8 = 16	-

 Copyright © Huawei Technologies Co., Ltd.

Field Name	Bit Width (Unit: Bit)	Description
VLAN	4 x 8 = 32	At most one VLAN tag is supported, and TPID can only be 0x8100.
IP header	12 x 8 = 96	An IP header consists of: IP version (4 bits) Header length (4 bits) Service type (8 bits) Total length (16 bits) Identifier (16 bits) Flag (3 bits) Fragment offset (13 bits) Lifetime (8 bits) Protocol (8 bits) Header checksum (16 bits)
SA_IP	4 x 8 = 32	Source IP Address. If an NE port is configured with an IP address, this field is filled with this IP address. If an NE port is not configured with an IP address, this field is filled with the NE's IP address.
DA_IP	4 x 8 = 32	Destination IP Address. When the destination IP address is not configured, the multicast encapsulation mode is used. This field is automatically filled with the destination IP address. It is filled with 224.0.0.107 for Pdelay packets and filled with 224.0.1.129 for other packets. When the destination IP address is configured, the unicast encapsulation mode is used. This field is filled with the configured destination IP address or with the SA_IP of received packets.
SPN	2 x 8 = 16	Source port ID. It is filled by the system with the ID of the NE port that transmits the packet.
DPN	2 x 8 = 16	Destination port ID, which is used to identify whether the packet is a 1588 packet. It is filled by the system.
UDP _Len	2 x 8 = 16	Data packet length. It is calculated by the system automatically using the following formula: UDP header length + 1588v2 payload length.

 Copyright © Huawei Technologies Co., Ltd.

Field Name	Bit Width (Unit: Bit)	Description
UDP_checksum	2 x 8 = 16	Header checksum. The value is 0 if the check is not required.
Payload	N = (A value ranging from 44 to 64) x 8	-
FCS	4 x 8 = 32	-

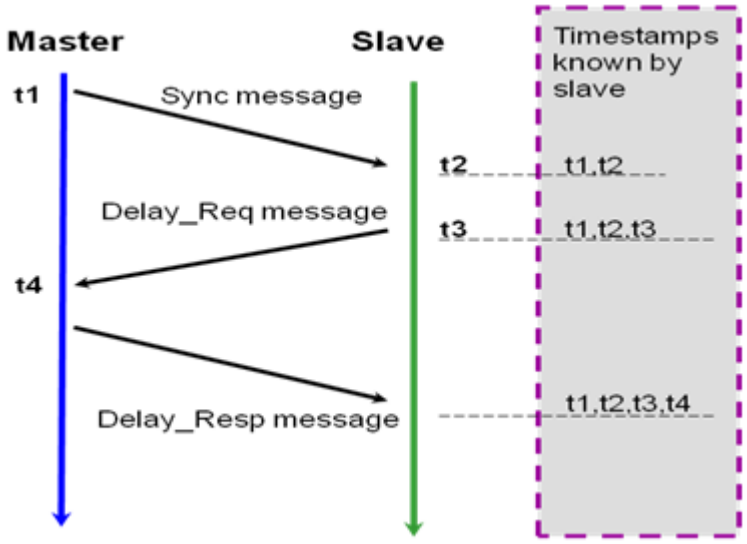
14.4.3 1588v2 Principle

14.4.3.1 Clock Synchronization Mechanism

Time Synchronization Principle

The master and slave exchanges IEEE 1588v2 messages in the following procedure:

Figure 14-13 IEEE 1588v2 time synchronization principle



1. The master sends a Sync message at t1 and carries the t1 timestamp in the Sync message.
2. The slave receives the Sync message at t2, locally generates the t2 timestamp, and extracts the t1 timestamp from the Sync message.

 Copyright © Huawei Technologies Co., Ltd.

3. The slave sends a Delay_Req message at t3 and locally generates the t3 timestamp.
4. The master receives the Delay_Req message at t4, locally generates the t4 timestamp, and sends the Delay_Req message with the t4 timestamp back to the slave.
5. The slave extracts the t4 timestamp from the Delay_Resp message after receiving it.

"Delayms" is the path delay in the direction from the master to the slave, "Delaysm" is the path delay in the direction from the slave to the master, and "Offset" is the time offset between the slave and master. All of them are variables.

t2-t1=Delayms+Offset

t4-t3=Delaysm-Offset

When Delayms = Delaysm, that is, when the transmit and receive links between the master and slave are symmetric, the following formula is satisfied:

Offset=[(t2-t1)-(t4-t3)]/2

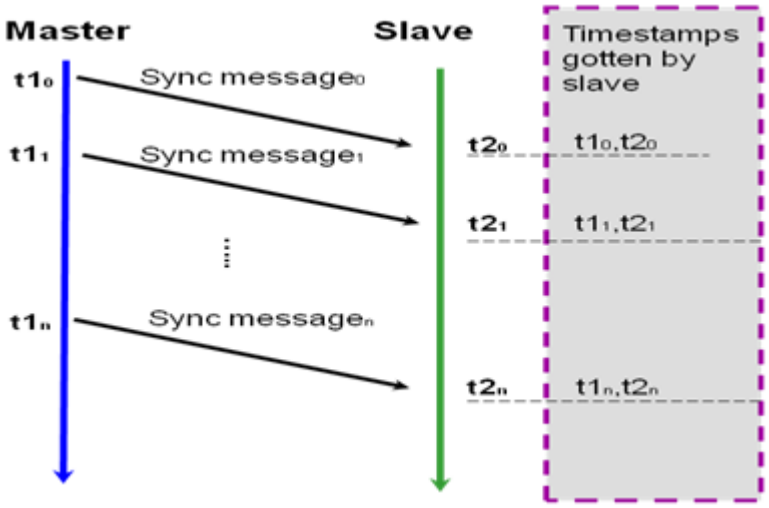
The slave can calculate the time offset between itself and the master based on the t1, t2, t3, and t4 timestamps and then corrects its own time to get synchronized with the master.

The preceding principle shows that IEEE 1588v2 time synchronization is based on the link symmetry between the master and slave. If the transmit and receive links are asymmetric, synchronization errors will occur and it will be half of the link delay asymmetry.

Frequency Synchronization Principle

The IEEE 1588v2 protocol implements frequency synchronization by exchanging Sync messages between the master and slave.

Figure 14-14 IEEE 1588v2 frequency synchronization principle



The master periodically sends Sync messages to the slave. If the slave frequency is synchronized to the master frequency, then the accumulative time errors within

 Copyright © Huawei Technologies Co., Ltd.

the same time periods are the same, as long as the path delay changes are neglected.

In other words, $t2_1$-$t2_0$=$t1_1$-$t1_0$, $t2_2$-$t2_1$=$t1_2$-$t1_1$, $t2_3$-$t2_2$=$t1_3$-$t1_2$,...$t2_n$-$t2_0$=$t1_n$-$t1_0$.

If $t2_n$-$t2_0$ is greater than $t1_n$-$t1_0$, then the slave frequency is higher than the master frequency, which means the slave frequency must be decreased. Reversely, the slave frequency must be increased.

14.4.3.2 BMC Algorithm

Background

The 1588v2 synchronization process involves two phases:

1. Establish the master/slave hierarchy.

   In a clock domain, ordinary clock (OC) nodes and boundary clock (BC) nodes run the best master clock (BMC) algorithm based on the information carried in Announce packets received from each port to determine the reference clock source of each OC node and BC node and the status of each port, and set up the master-slave hierarchy and tracing path of the entire synchronization network.
2. Frequency/time synchronization

   After the master/slave hierarchy and tracing paths are established, the master and slave exchange event messages such as Sync, Delay_Req, and Delay_Resp to implement frequency or time synchronization.

This section describes the BMC algorithm principles.

Overview

The BMC algorithm, defined in the 1588v2 protocol, is used to determine the master-slave relationship among clocks in a network. With this algorithm, clocks in the network are classified into master clock and slave clock. Slave clocks lock the frequency or time of the master clock. In the event of changes in the network or clock source attributes, the best master clock is re-selected using the BMC algorithm so that the time and clock in the entire network are synchronized.

- From the network aspect, the BMC algorithm helps form a tree structured (Grandmaster as the root and best clock source) master-slave synchronization hierarchy for clocks in the entire network.
- From the node aspect, the BMC algorithm helps determine the master clock in each clock node. All Announce messages received on the ports of a given clock node and the local clock are compared to determine which clock is the best.

If the best clock is an external clock, the external clock is determined as the master clock, that is, the clock is traced. If the best clock is the local clock, the local clock is the Grandmaster clock.

Algorithm Implementation

The BMC algorithm involves data set comparison, status decision, state machine, and data set update. The data set members include priority, clock quality (level,

 Copyright © Huawei Technologies Co., Ltd.

precision, and offset), and stepRemoved (distance to the Grandmaster). These members are contained in ANNOUNCE packets defined by the protocol. An NE obtains the information about these members from the ANNOUNCE packets received on its ports. The NE uses the data set comparison algorithm to select the best external clock source (Ebest). Then the NE uses the status decision and state machine to determine the master/slave state of each port based on the Ebest, local clock D0, and data sets on the ports (Erbest). If the NE's clock is the master clock on the entire network, the NE's all ports transmit clock signals to other NEs. If the NE's clock is the slave clock, a port on the NE must be the slave port for tracing the master clock. The NE transmits the obtained master clock data to lower-layer NEs in ANNOUNCE packets through its master port.

Data set comparison algorithm

The algorithm compares two groups of data that represent two clock ports. One group of data represents the default features of the local clock, and the other group of data represents the information contained in the synchronization packet received from an external port. The algorithm compares the following attributes in the data sets in order to obtain the best ANNOUNCE packet:

- priority1: indicates the priority configured by a user. It identifies a clock's relative priority in the master clock set.
- clockClass: indicates the clock source's level.
- clockAccuracy: indicates the clock's precision.
- OffsetScaledLogVariance: indicates the clock's stability. It is the quality of clock signals that are received by an NE. The NE obtains the value using Allan deviation.
- priority2: indicates the priority configured by a user. It is used to select a better clock source when other attributes of clock sources are the same.
- Clock Identity: indicates the clock ID on a device.

After the preceding priority-based sorting is complete, when two interfaces on the same NE receive two Announce messages indicating the same external master clock and the comparison results are the same, distance between the local clock and the external master clock (indicated by the stepsRemoved field in the Announce message) is used to determine the clock based on the number of hops on the link. (The clock received by the port with fewer hops to the Grandmaster clock source is selected.) This situation usually occurs in a PTP system with a loop.

Status decision algorithm

The status decision algorithm involves the data sets D0, Erbest, and Ebest. The algorithm determines the status decision code and recommended status for each port based on the data set comparison results. The recommended status is a factor for triggering the state machine. The state machine obtains a port's next status based on the triggering factor and current status. The status decision code is used to select a data set to update the global date set of the NE.

State machine

The state machine obtains a port's next status based on the triggering factor and current status. The trigger condition may be the recommended status of the port or some external events that may cause the status change, such as port enabling/disabling (DISABLE/ENABLE), port going online/offline (PORT LINKUP/DOWN), and timeout (TIMEOUT) interruption.

 Copyright © Huawei Technologies Co., Ltd.

Data set update

An NE uses the status decision code of each port to determine whether to use Ebest or D0 to update the global data set. The updated global data set is used as the tracing data set of the NE and is transmitted by the NE's master port to lower-layer NEs.

In the LTE carrying scenario, the BMC uses only the data set comparison algorithm. The algorithm is implemented as follows: Each port uses the algorithm to compare the ANNOUNCE packets that contain clock information and select the best clock source (Erbest) received on the port.

BMC Implementation in an FTTM LTE Scenario

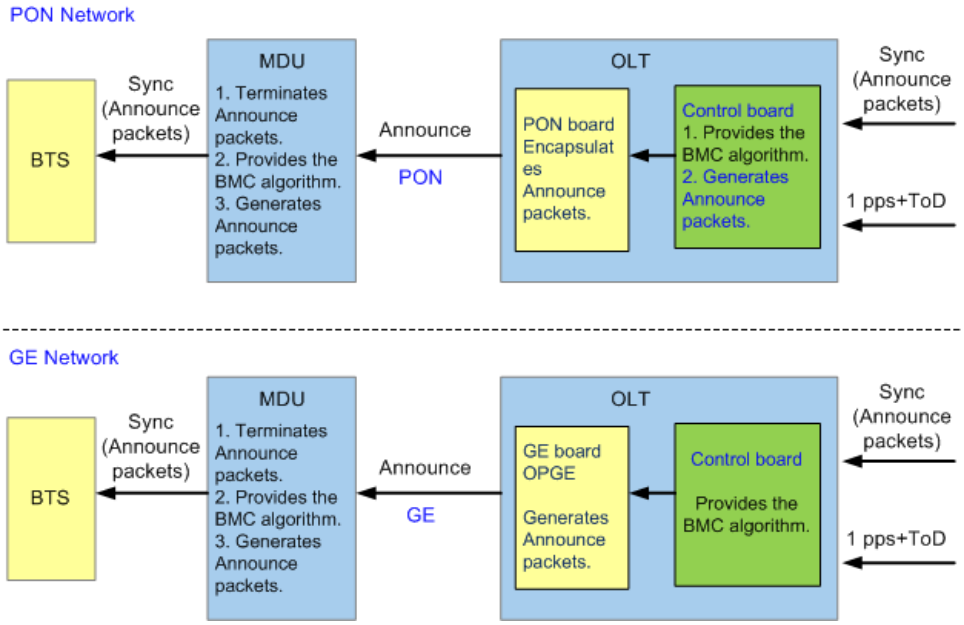
As shown in **Figure 14-15**, in the PON networking scenario carried by FTTM LTE base stations, the overall BMC solution is as follows:

1. An OLT obtains ANNOUNCE packets from an uplink Ethernet port.
2. The control board of the OLT obtains the time source information from the Announce message, compares the time source with other synchronization sources in the system through the BMC algorithm, and selects the best clock as the system time source.
3. The OLT constructs Announce packets and broadcasts the packets to downstream MDUs through the GEM ports.
4. The MDU obtains the Announce message from the PON line, obtains the time source information from the Announce message, compares the time source with other synchronization sources in the system through the BMC algorithm, and selects the best clock as the system time source.
5. An MDU constructs ANNOUNCE packets and transmits them to lower-layer base stations using user-side ports.

The difference between the GE networking scenario and the PON networking scenario is as follows: The OLT constructs the Announce packets and then issues the PTP packets to the downstream MDU through the GE.

 Copyright © Huawei Technologies Co., Ltd.

Figure 14-15 BMC implementation in an FTTM LTE scenario



14.4.3.3 Delay Compensation Mechanism

Context

Based on the principles described, 1588v2 time synchronization is based on delay symmetry on transmit and receive links between the master and slave. If the link delay between the master and slave is asymmetric, synchronization values will be inaccurate, and the error will be half of the asymmetric link delay.

The transmission delay on a one-meter fiber is about 5 ns, so asymmetry on a one-meter fiber causes a 2.5 ns time synchronization error, and asymmetry on a 400-meter fiber causes a 1 us error. On a practical network, it is difficult to accurately control the E2E fiber asymmetry on the entire network within 400 m. For TD-SCDMA and LTE-TDD which require a synchronization precision of ±1.5 us, a 1 us error is obviously intolerable. Therefore, in deployment of the 1588v2 time synchronous network, fiber asymmetry in the network must be seriously considered.

Solution

Currently, three solutions to fiber asymmetry are used in the industry:

- Point-by-point measurement by instrument: During deployment and maintenance, a device's 1588 time synchronization precision is measured by an 1588v2 instrument point by point. Fiber asymmetry is calculated based on the measured precision, and compensation is made accordingly. Alternatively, an OTDR can be used to measure asymmetry of each pair of fibers, and compensation can be made. The OTDR test interrupts services and the 1588v2 synchronization result cannot be viewed directly. Therefore, the 1588v2 meter is recommended for the test.
- Bidirectional transmission over a single fiber: A device uses a bidirectional single-fiber optical module to support signal transmission and reception on a

 Copyright © Huawei Technologies Co., Ltd.

single fiber. This solution substantially resolves the asymmetry issue on dual-fiber links, and no measurement or compensation is required. In this solution, bidirectional link delay due to different wavelengths used for transmitting and receiving signals is generally small and can be automatically calculated and compensated by devices. According to G.652, the time error introduced by FE single-fiber bidirectional transmission is about 1.06 ns per kilometer, and the time error introduced by GE single-fiber bidirectional transmission is about 0.544 ns per kilometer, which is far less than the time error of 2.5 ns per meter for common dual-fiber transmission. After the single-fiber bidirectional optical module is used, you do not need to perform measurement onsite during deployment or maintenance.

- Automatic ring network measurement: A device uses a common dual-fiber optical module. During deployment, an instrument measures and compensates for asymmetry. During maintenance, asymmetry can be automatically measured on the ring network to avoid the necessity for onsite measurement upon fiber disconnection. This solution reduces the workload for repairing fiber disconnection. This solution utilizes the BMC algorithm in 1588v2. When fiber disconnection occurs on a ring network, services are automatically switched to the standby link, and the time synchronization precision of the nodes on the ring network stays in the usable range. When the faulty fiber is re-connected, the devices first automatically calculate and report asymmetry on the new fiber link and then compensate for asymmetry on the NMS. After that, services are switched back to the active link.

Delay Compensation Mode

Delay compensation can be length compensation or time compensation.

The relationship between the length compensation and time compensation is as follows: the delay compensated for each 1 m twisted pair is 8 ns, and the delay compensated for each 1 m coaxial cable is 4.5 ns.

14.4.4 1588v2 Network Application

14.4.4.1 Recommended: Synchronization Application (Network-wide 1588v2 and Synchronous Ethernet Deployment)

When bearer network devices support 1588v2 synchronization, time signals are injected from the bearer network.

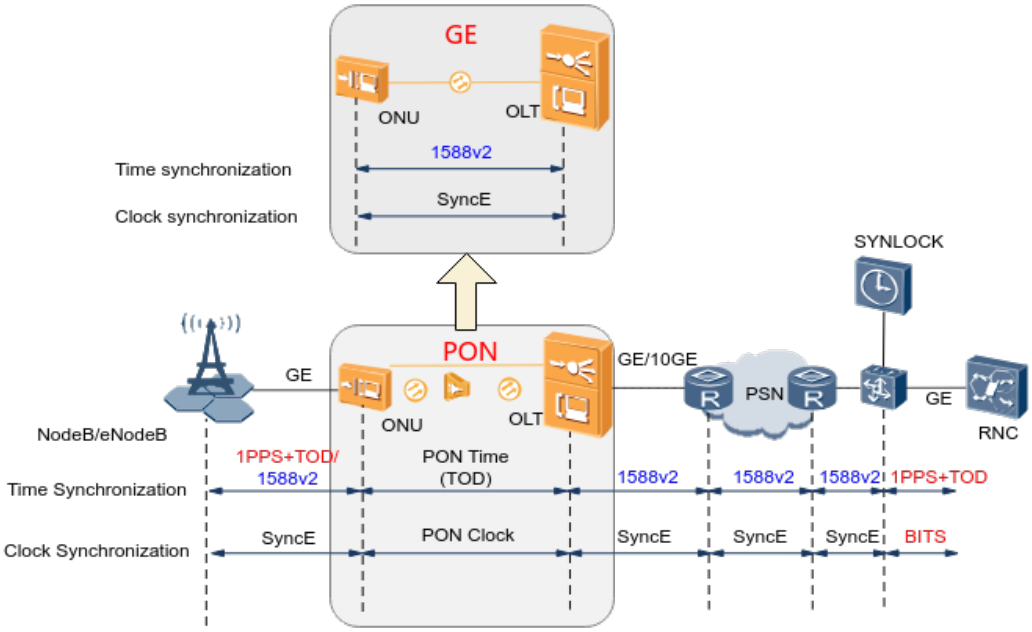
- Time is synchronized in 1588v2 mode.
- When devices on the entire network support synchronous Ethernet (SyncE), SyncE is recommended for clock synchronization, as follows.

NOTE

SyncE uses Ethernet bit streams to recover clock signals for Ethernet clock synchronization. The implementation mode is similar to that on an SDH/PDH network. SyncE uses high-precision clock signals as the transit reference in the transmit direction. It restores and extracts the clock signals at the receive end. The physical layer transmits and receives the clock signals and is compatible with traditional networks.

 Copyright © Huawei Technologies Co., Ltd.

Figure 14-16 Synchronization application (network-wide 1588v2 and SyncE deployment)



Time Synchronization

- Bearer network devices are used as an Ordinary Clock (OC) device. A bearer network device receives time signals through a 1PPS+TOD port and uses 1588v2 to synchronize time signals with other bearer network devices.
- Access devices are regarded as a boundary clock (BC) device as a whole. An OLT transmits 1588v2 time signals through a GE or 10GE upstream link to ONUs, and the ONUs output 1588v2 time to downstream devices through user-side GE links. Between the OLT and an ONU, time is transmitted using PON links in ONT Management and Control Interface (OMCI) mode or using GE links in PTP packet mode.
- Base stations are used as an Ordinary Clock (OC) device.
  - If a base station does not support 1588v2 synchronization, the base station uses a 1PPS+TOD port to receive time signals.
  - If a base station supports 1588v2 synchronization, the base station uses 1588v2 to synchronize time signals.

Clock Synchronization

- Bearer network devices use SyncE to synchronize clock signals with each other.
- An OLT transmits SyncE through a GE or 10GE upstream link to ONUs, and the ONUs output SyncE to downstream devices through user-side GE links. Between the OLT and an ONU, time is transmitted using PON or GE links.
- The base stations use SyncE to synchronize clock signals.

14.4.4.2 Synchronization Application (Network-wide 1588v2 Deployment)

If devices at the bearer network support 1588v2 synchronization, time source is injected from the devices, and time synchronization is implemented using 1588v2.

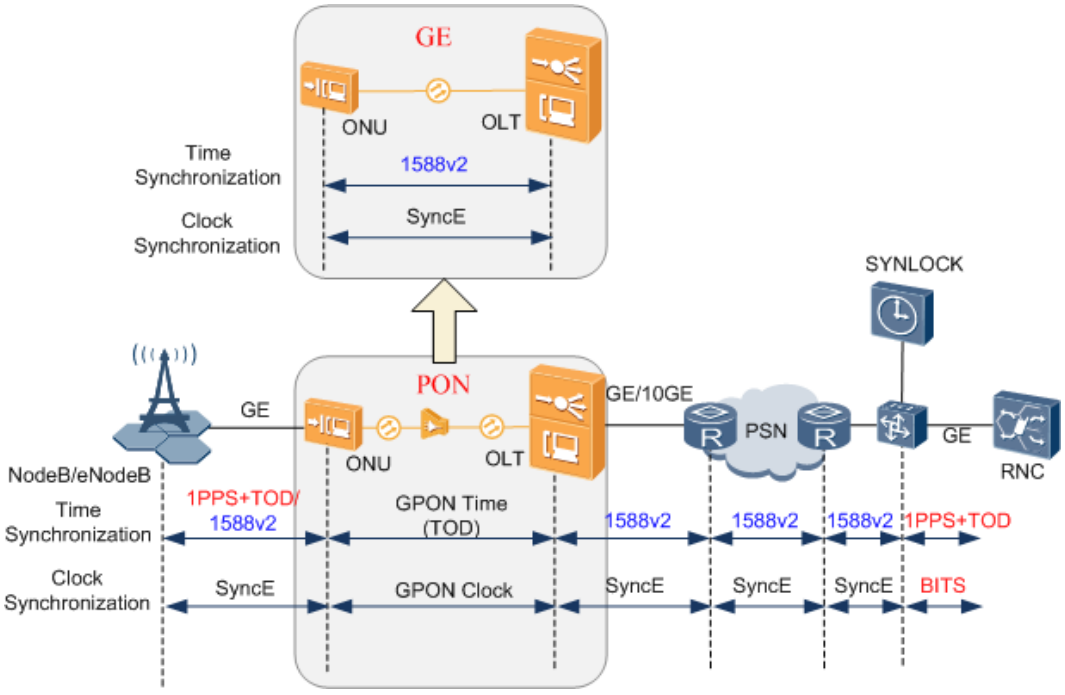
 Copyright © Huawei Technologies Co., Ltd.

If devices at the entire network do not support Synchronous Ethernet (SyncE), clock synchronization is implemented using 1588v2.

Time synchronization must be implemented on 3G, 4G, or Long Term Evolution (LTE) base stations and can be provided by the bearer network. When bearer network devices support 1588v2 synchronization, time signals are injected to a bearer network device and be synchronized, as shown in **Figure 14-17**.

- Bearer network devices are used as an Ordinary Clock (OC) device to inject the synchronization time source using the 1PPS+TOD port. A bearer network device receives time signals through a 1PPS+TOD port and uses 1588v2 to synchronize time and clock signals with other bearer network devices.
- OLTs and ONUs are regarded as a boundary clock (BC) device as a whole. An OLT transmits 1588v2 time signals through a GE or 10GE upstream link to ONUs, and the ONUs output 1588v2 time to downstream devices through user-side GE links. Between the OLT and an ONU, time and clock signals are transmitted using PON links in ONT Management and Control Interface (OMCI) mode or using GE links in PTP packet mode.
- Base stations are used as an Ordinary Clock (OC) device.
  - If a base station does not support 1588v2 synchronization, the base station uses a 1PPS+TOD port to receive time signals, and use SyncE to synchronize clock signals.
  - If a base station supports 1588v2 synchronization, the base station uses 1588v2 to synchronize time and clock signals.

Figure 14-17 Synchronization application (network-wide 1588v2 deployment)



14.4.4.3 Synchronization Application (Clock or Time Signal Injection from an OLT)

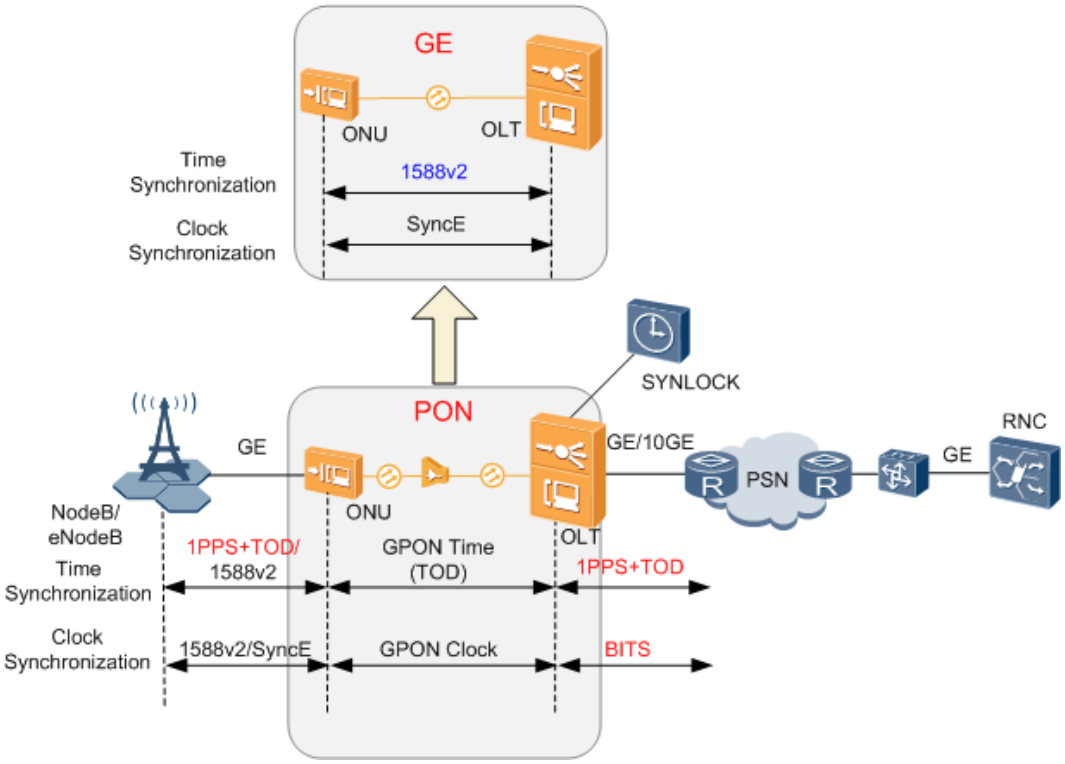
Devices at the bearer network do not support 1588v2 synchronization, and time source and clock source are injected from an OLT.

 Copyright © Huawei Technologies Co., Ltd.

3G and 4G long term evolution (LTE) base stations require time synchronization. When devices on the bearer network do not support 1588v2 synchronization, time signals can be injected from an OLT, as shown in **Figure 14-18**.

- Time signals are injected from an OLT using a 1PPS+TOD port. The BITS synchronization clock source is injected into the building integrated timing supply (BITS) port.
- OLTs and ONUs are regarded as a boundary clock (BC) device as a whole. ONUs output 1588v2 time to downstream devices through user-side GE links. Between the OLT and an ONU, time and clock signals are transmitted using PON links in ONT Management and Control Interface (OMCI) mode or using GE links in PTP packet mode.
- Base stations are used as an Ordinary Clock (OC) device.
  - If a base station does not support 1588v2 synchronization, the base station uses a 1PPS+TOD port to receive time signals, and use SyncE to synchronize clock signals.
  - If a base station supports 1588v2 synchronization, the base station uses 1588v2 to synchronize time and clock signals.

Figure 14-18 Synchronization application (1588v2 clock or time signal injection from an OLT)



14.4.4.4 Time and Clock Synchronization Protection Application

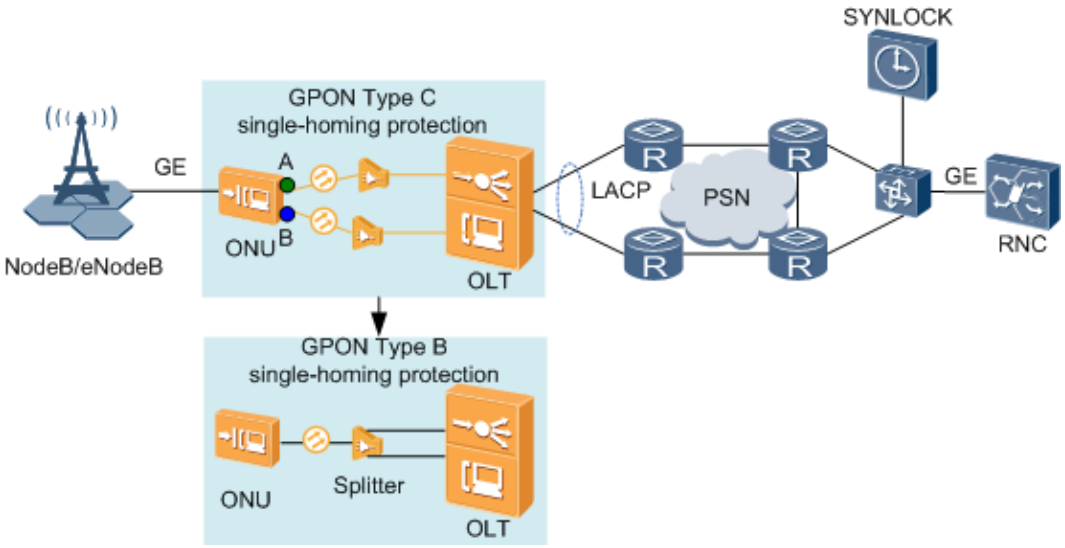
A protection solution is provided on the bearer network and multiple time sources and clock sources are available. The time and clock synchronization protection feature improves network security.

 Copyright © Huawei Technologies Co., Ltd.

As shown in **Figure 14-19**, GPON Type C/Type B single-homing protection is configured between the ONU and two OLTs, and LACP is configured between the OLTs and bearer network devices.

- When the working link becomes faulty, protection switching is triggered and data is switched to the standby link. This protection mechanism ensures uninterrupted data transmission.
- As shown in **Figure 14-19**, port A is assumed to be a master port and port B is a slave port. If the active link is faulty, the clock source must also be switched to the standby link. That is, the link is switched from port A to port B. This prevents signals from being lost.
- After the faulty link restores, the system will not immediately switch the clock source back to the original active link and the clock source of port B is preferentially traced.

Figure 14-19 GPON Type C dual-homing protection



NOTE

Other protection features can be deployed between the OLT and ONUs, or between the OLT and the bearer network. For details about the protection features, see feature description.

 Copyright © Huawei Technologies Co., Ltd.

Table 14-8 Time and clock source selection modes on an ONU/OLT

Synchronization Type	Time and Clock Source Selection Modes on an ONU/OLT
Time synchronization	• Static selection based on priorities configured on an ONU/OLT Users must configure a unique priority for each external time source. The system selects the time source that has the highest priority and is in the normal state as the reference time. • Automatic selection based on best master clock (BMC) The BMC algorithm classifies clocks on a network into a master clock and slave clocks. Slave clocks trace the frequency or time of the master clock. When the network topology or the attribute of a clock source changes, the algorithm selects a new grandmaster clock so that clocks and times are synchronized on the entire network. For details, see **14.4.3.2 BMC Algorithm**.
Clock synchronization	• Static selection based on priorities configured on an ONU/OLT Users must configure a unique priority for each external clock source. The system selects the clock source that has the highest priority and is in the normal state as the reference clock. • Automatic selection based on Synchronization Status Message (SSM) levels The system obtains the clock quality level from received clock source signals (users must manually configure the quality level for a clock source whose SSM level cannot be obtained). Then the system automatically selects a clock source as the reference clock when the source meets the following three conditions: has the highest level; the level is not smaller than the lower SSM level; is in the normal state. If the two external clock sources have the highest SSM quality level, the system selects a clock source with the highest priority as the system reference clock.

14.4.5 Configuring the 1588v2 Function

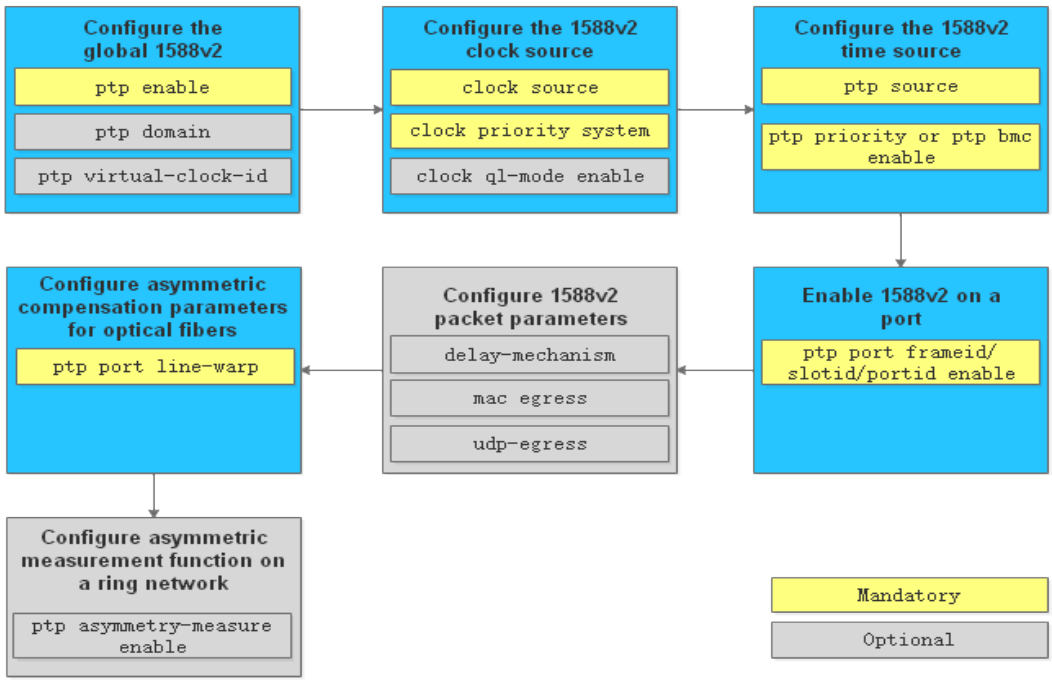
3G and 4G long term evolution (LTE) base stations require frequency (clock) and phase (time) synchronization. This section describes how to configure 1588v2 to implement clock synchronization and sub-microsecond-level time synchronization.

Configuration Process

Figure 14-20 shows the process of configuring the 1588v2 function.

 Copyright © Huawei Technologies Co., Ltd.

Figure 14-20 Process of configuring the 1588v2 function



Procedure

Step 1 Configure the global 1588v2 function.

1. Run the **ptp enable** command to enable the 1588v2 function on a device.
2. **Optional:** Run the **ptp domain** command to specify a 1588v2 time domain.

   After the configuration, a time domain has a clock source. All devices in the time domain use the clock source. Ensure that the MDU time domain is the same as that of the OLT and upper-layer bearer network device.
3. **Optional:** Run the **ptp virtual-clock-id** command to configure the 1588v2 virtual clock ID of a device.

   The clock ID uniquely identifies a clock node in a 1588v2 domain.

   If this parameter is not configured, its value consists of the organizational unique identifier (OUI)+product ID+last four bytes of the MAC address. The OUI is 00-25-9e, and the product ID is 40. For example, if the MAC address of the system is 00e0-fc00-a025, the corresponding default clock ID is 00259e405898a025.

Step 2 Configure a 1588v2 clock source. After the configuration, device clocks on the network are synchronized.

1. Run the **clock source** command to configure a 1588v2 clock source.

   The clock source can be the synchronous Ethernet clock or 1588v2 clock recovered from a GE or 10GE line on the network side or the BITS clock.

   If the network supports the synchronous Ethernet clock, the clock source preferentially uses the synchronous Ethernet clock.
2. Run the **clock priority system** command to set the priority of the clock source. A device may trace multiple clock sources for synchronizing the clock. It selects the clock source with the highest priority as the clock source.

 Copyright © Huawei Technologies Co., Ltd.

p0 indicates the highest priority, and p9 indicates the lowest priority.

3. **Optional:** Configure the clock source selection mode when the protection mechanism is configured.

   A clock source can be statically or dynamically selected.

   - To statically select a clock source based on the local priority, perform **Step 2.2** to set a local priority for a clock source. The device selects the clock source with the highest priority as the clock source.
   - To automatically select a clock source using the Synchronization Status Message (SSM):
     i. Run the **clock ql-mode enable** command to configure the status of the SSM source selection function.
     ii. Run **clock ql** command to configure the SSM quality level of the clock source. The device selects a clock source as the clock source when the clock source has the optimal quality level that is greater than or equal to the lowest SSM level of the clock input source and is running properly.

Step 3 Configure a 1588v2 time source. After the configuration, device time on the network is synchronized.

1. Run the **ptp source** command to configure a 1588v2 time source.

   The time source can be the 1588v2 time recovered from a GE or 10 GE line on the network side or BITS time. If the network does not support 1588v2 time synchronization, the time source can use the BITS time.

2. **Optional:** When the BITS time is used, run the **ptp bits** command to configure the attributes of a BITS clock source, such as the precision and class of the clock source, and the priority of the clock signal.
3. Select a time source when the protection mechanism is used.

   A time source can be statically or dynamically selected.

   - To statically select a time source based on the local priority, run the **ptp priority** command to set the priority of the time source.
   - To dynamically select a time source based on the best master clock (BMC), run the **ptp bmc enable** command to enable the automatic BMC source selection function.

Step 4 Enable 1588v2 on a port to transmit 1588v2 packets.

Run the **ptp port** to enable 1588v2 on an Ethernet port.

Enable 1588v2 on all Ethernet ports that need to transmit 1588v2 packets.

Step 5 **Optional:** (Optional) Configure 1588v2 parameters.

Set the timestamp mode of the 1588v2 packets, MAC encapsulation mode, and UDP encapsulation mode.

1. Run the **ptp port delay-mechanism** command to configure the delay measurement mode for 1588v2 ports. The default mode is delay.

   The delay measurement mode on ports of all the devices in a 1588v2 synchronization network must be the same.

 Copyright © Huawei Technologies Co., Ltd.

2. Run the **ptp port frameid/slotid/portid clock-step { one-step | two-step }** command to configure the step mode for 1588v2 ports.

   By default, 1588v2 packets carry timestamps in one-step mode. The 1588v2 ports identify Follow_Up packets in two-step mode in the Rx direction for communicating with other products.

3. Run the **ptp port frameid/slotid/portid interval { announce value | req value | sync value }** command to configure the interval at which 1588v2 packets are sent.
   - 1588v2 frequency synchronization:

     The peer-end master device must send Sync packets at an interval longer than or equal to 32 packets per second.
   - 1588v2 time synchronization:
     - The peer-end master device must send Sync packets at an interval longer than or equal to 1 packet per second.
     - The local-end device must send Delay_request or Pdelay_request packets at an interval longer than or equal to 1 packet per second.

4. **Optional:** Configure the encapsulation mode for 1588v2 packets.
   - Run the **ptp port mac-egress** command to configure the MAC encapsulation mode for 1588v2 packets to be forwarded by the port. The packets are forwarded in Layer 2 mode.
     - When you specify **destination-mac**, the 1588v2 packets are forwarded in Layer 2 unicast mode.
     - When you do not specify **destination-mac** *destination-mac-value*, the 1588v2 packets are forwarded in Layer 2 multicast mode.
   - Run the **ptp port frameid/slotid/portid udp-egress { [destination-mac destination-mac-value] | [source-ip source-ip-addr [destination-ip destination-ip-addr][dscp dscp-value][vlan vlan-id [priority priority-value]]] }** command to configure the UDP encapsulation mode for 1588v2 packets to be forwarded on the port. The packets are forwarded in Layer 3 mode.
     - When you specify **destination-ip** *destination-ip-addr*, the 1588v2 packets are forwarded in Layer 3 unicast mode.
     - When you do not specify **destination-ip** *destination-ip-addr*, the 1588v2 packets are forwarded in Layer 3 multicast mode.

Step 6 Configure asymmetric compensation parameters for optical fibers.

Run the **ptp port** frameid/slotid/portid line-warp linewarptype linewarpdir linewarpvalue command to configure asymmetric compensation parameters for optical fibers.

Asymmetric optical fibers between two devices result in time difference for data transmission and reception. During site deployment, measure and calculate the time differences in the two directions and compensate the time on the devices by configuring cable transmission deviations.

Step 7 **Optional:** Configure the asymmetric measurement function on a ring network.

 Copyright © Huawei Technologies Co., Ltd.

Run the **ptp asymmetry-measure enable** command to configure the asymmetric measurement function on a ring network.

This function utilizes the BMC algorithm in IEEE 1588v2. When fiber disconnection occurs on the ring network, services are automatically switched to the standby link, and the time synchronization precision of the nodes on the ring network stays in the usable range. When the faulty fiber is re-connected, the devices first automatically calculate and report asymmetry on the new fiber link and then compensate for asymmetry on the NMS. After that, services are switched back to the active link.

----End

Result

- Run the **display clock source** command to check whether the clock source in the system is 1588v2 and whether it is running properly. In normal cases, **State** is **successful**.
- Run the **display ptp source** command to check whether the clock source in the system is 1588v2 and whether it is running properly. In normal cases, **Lock state** is **successful**.

When clock tracing is abnormal, rectify the fault based on the handing suggestions in the reported alarm or event.

Example

An MDU connects to a 4G LTM eNodeB using a GE port. The MDU connects to the OLT using the GPON uplink port and then to the radio network controller (RNC) over the upper-layer network to carry 4G services over the access network. The 4G LTE eNodeB requires high precision time synchronization and the 1588v2 time is deployed on the network.

The parameters planned on the OLT are as follows:

- Time domain: 1
- Clock or time source input: GE uplink port 0/8/0
- Clock or time source output: GPON service port 0/3/1
- Packet encapsulation mode: MAC (default mode) (packets are forwarded in Layer 2 multicast mode)

The parameters planned on the MDU are as follows:

- Time domain: 1
- Clock or time source input: GPON uplink port 0/0/0
- Clock or time source output: GE service port 0/1/1
- Packet encapsulation mode: MAC (default mode) (packets are forwarded in Layer 2 multicast mode)

Configurations on the OLT side

```
huawei(config)#ptp enable
huawei(config)#ptp domain 1
huawei(config)#clock source 0 1588
huawei(config)#clock priority system 0
huawei(config)#ptp source 0 0/8/0
huawei(config)#ptp bmc enable
```

 Copyright © Huawei Technologies Co., Ltd.

Configurations on the MDU

```
huawei(config)#ptp enable
huawei(config)#ptp domain 1
huawei(config)#clock source 0 0/0/0
huawei(config)#clock priority system 0
huawei(config)#ptp source 0 0/0/0
huawei(config)#ptp port 0/1/1 enable
huawei(config)#ptp bmc enable
```

14.4.6 Configuring 1588v2-related Delay Compensation for Asymmetric Fibers

Prerequisites

A global positioning system (GPS) and clock signal tester, such as a Paragon tester, are available.

The 1588v2 precision of the upper-layer device meets usage requirements.

1588v2-related Delay Compensation Rules for Asymmetric Fibers

- After the first compensation on a network, if the 1588v2 precision meets clock precision requirements (1 us) of a base station, no more compensation is required for this base station.
- If the 1588v2 precision is required to improve (to be within 500 ns) for a base station, or the 1588v2 precision of some base stations fails to meet clock precision requirements after the compensation on the entire network, a second compensation is required for such base stations.
- If the capacity of a network needs to be expanded (for example, a ring is added), and the compensation on new base stations or sites must not change the original compensation data, separate compensation is performed on the new base stations or sites.

Context

An OLT can connect to ONUs in PON or GE mode:

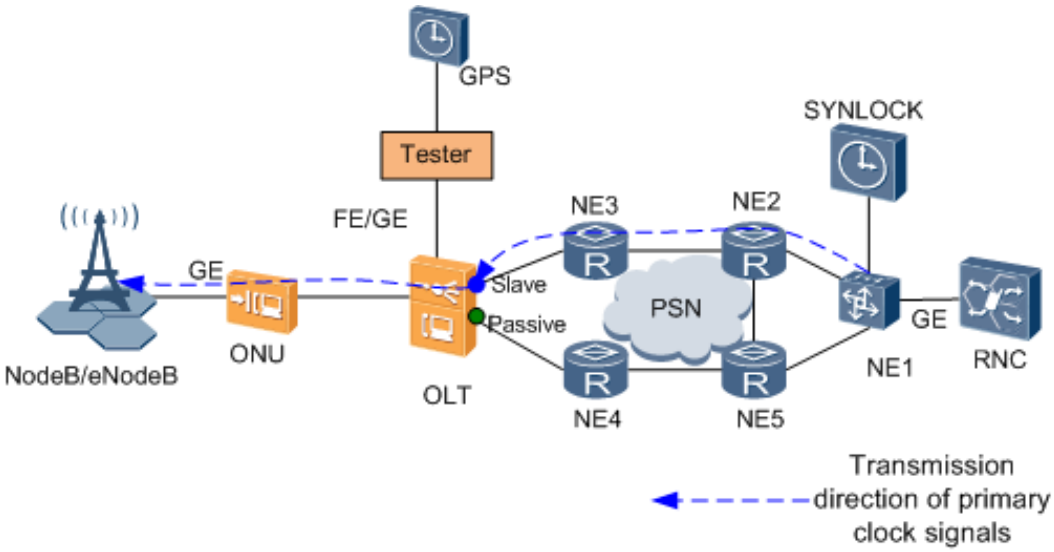
- In PON connections, the optical fibers between the OLT and ONUs are symmetric. Therefore, no asymmetric delay compensation is required for the optical fibers. In this case, you only need to measure the OLT and implement delay compensation on the optical fibers between the OLT and its upper-layer device.
- In GE connections, the optical fibers between the OLT and ONUs may be asymmetric. In this case, you need to measure the ONUs and implement delay compensation on the optical fibers not only between the OLT and its upper-layer device but also between the OLT and ONUs.

As shown in the following figure, the OLT is used as a test node.

- The OLT connects to a tester using an FE/GE/10GE port.
- The slave port on the OLT connects to NE 3 for receiving primary clock signals.
- The passive port on the OLT connects to NE 4 for receiving standby clock signals.

 Copyright © Huawei Technologies Co., Ltd.

Figure 14-21 Delay compensation networking



Procedure

Step 1 Perform delay compensation on the slave port of the OLT.

1. Calculate the difference between the data output by the FE or GE port after the OLT traces the 1588v2 clock and the data measured by the GPS. Record the difference as D1.
2. Check whether D1 is within ±20 ns. If it does, go to **Step 2.1**. If it does not, go to **Step 1.3**.
3. Run the **ptp port** *frameid/slotid/portid* **line-warp** *linewarptype linewarpdir linewarpvalue* command to set the compensation value of the slave port to d, which is equal to D1.
4. Calculate the difference between the output 1 PPS data after the OLT traces the 1588v2 clock and the data measured by the GPS. Record the difference as D2.
5. Check whether D2 is within ±20 ns. If it does, go to **Step 2.1**. If it does not, go to **Step 1.3**.

Step 2 Perform delay compensation on the passive port of the OLT.

1. Check whether the OLT provides a passive port. If it does, go to **Step 3**. If it does not, go to **Step 2.2**.
2. Run the **ptp port** *frameid/slotid/portid* **disable** command to disable 1588v2 on the original slave port.
3. Check whether the passive port has been switched to the slave port. If it does, go to **Step 2.4**.
4. Calculate the difference between the output 1 PPS data after the OLT traces the 1588v2 clock and the data measured by the GPS. Record the difference as D3.
5. Check whether D3 is within ±20 ns. If it does, go to **Step 3**. If it does not, go to **Step 2.6**.
6. Run the **ptp port** *frameid/slotid/portid* **line-warp** *linewarptype linewarpdir linewarpvalue* command to set the compensation value of the slave port to d, which is equal to D3.

 Copyright © Huawei Technologies Co., Ltd.

7. Calculate the difference between the output 1 PPS data after the OLT traces the 1588v2 clock and the data measured by the GPS. Record the difference as D4.
8. Check whether D4 is within ±20 ns. If it does, go to **Step 2.9**. If it does not, go to **Step 2.6**.
9. Enable 1588v2 on the original slave port and check whether this port has been switched to the status queried in **Step 2.1**.

Step 3 Perform delay compensation on the uplink port of an ONU.

Connect a tester to the ONU and perform delay compensation on the optical fiber between the OLT and this ONU. For details, see **Step 1**.

----End

14.4.7 1588v2 Maintenance and Diagnosis

This describes how to query the status of common alarms and how to handle the alarms.

The 1588v2 clock or time source tracing status can be verified by running the following command:

- **display clock source**: used to check whether the system reference clock source is a 1588v2 clock source and whether the clock source is functional.
- **display ptp source**: used to check whether the system reference time source is a 1588v2 time source and whether the time source is functional.
- **display ptp info**: used to check whether the time locking status and 1588v2 running data are correct.
- **display ptp timestamps**: used to query the timestamp information about the 1588v2 clock source.
- **display ptp packet-discard-reason port**: used to query the historical records of 1588v2 packet abnormality causes for a 1588v2 device port to know the abnormalities of 1588v2 packets on this port.

If any alarm or event is reported, rectify the fault according to the "Alarm and Event Reference".

14.4.8 1588v2 (based on IEEE) Reference Standards and Protocols

Document	Description
IEEE 1588-2008	Precision Clock Synchronization Protocol for Networked Measurement and Control Systems
ITU-T G.813	Timing requirements of SDH equipment slave clocks (SEC)
ITU-T G.823	The control of jitter and wander within digital networks which are based on the 2048 kbit/s hierarchy

 Copyright © Huawei Technologies Co., Ltd.

Document	Description
ITU-T G.8260	Definitions and terminology for synchronization in packet networks
ITU-T G.8261	Timing and Synchronization aspects in Packet Networks
ITU-T G.8262	Timing characteristics of Synchronous Ethernet Equipment slave clock (EEC)
G.987.3	XG-PON TC Layer Specification
G.988	ONU management and control interface specification (OMCI)

14.5 1588v2 (Based on G.8275.1)

G.8275.1 defines precision time protocol telecom profile for phase/time synchronization with full timing support from the network based on the IEEE 1588v2 precision time synchronization protocol.

1588v2 (based on G.8275.1) and 1588v2 (based on IEEE) have the same network applications. For details, see **14.4.4 1588v2 Network Application**.

14.5.1 Basic Concept

Packet Type

G.8275.1 supports packets of types Sync, Follow_Up, Announce, Delay_Req, and Delay_Resp.

- Packets Sync, Delay_Req, Follow_Up, and Delay_Resp are used to generate and exchange timing packets that are required when the delay mechanism is used for the synchronization between the ordinary clock (OC) and boundary clock (BC).
- Message Announce is used to set up a synchronization system, that is, the information related to best master clock algorithm (BMCA) is carried in the message Announce.

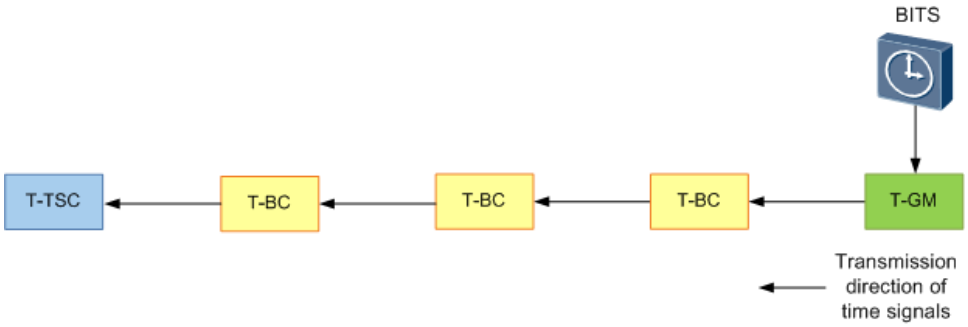
Device Model

The G.8275.1-supported devices are as follows:

- Telecom grandmaster (T-GM): The clock that can only be the master one supports one or more precision time protocol (PTP) ports, but cannot trace other PTP clocks.
- Telecom boundary clock (T-BC): The boundary clock that can be the master clock and it can trace other PTP clocks.
- Telecom time slave clock (T-TSC): The clock that can only be the slave clock.

The OLT or MA5694S can serve as a T-BC device.

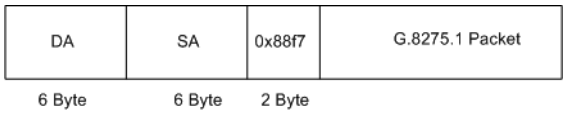
 Copyright © Huawei Technologies Co., Ltd.



Packet Encapsulation Mode

G.8275.1-defined packets are encapsulated in Layer 2 multicast mode (without tags), as shown in the following figure.

In the figure, EtherType is 0x88F7. There are two types of destination multicast MAC addresses: unforwardable multicast MAC address and forwardable multicast MAC address.



14.5.2 BMCA Algorithm

The G.8275.1 profile uses best master clock algorithm (BMCA) that is different from the clock source algorithm defined in 1588v2. Differences are:

1. BMCA defines the Boolean attribute of every port as notSlave.
   - If the port attribute is notSlave, the port is in the Master state and cannot be the Slave state.
   - If the port attribute is not notSlave, the port can be in the Slave state.
2. BMCA supports multiple Grandmaster clocks and one PTP domain traces one Grandmaster clock.
3. The port supports attribute LocalPriority to set the datasets Erbest and Ebest for port status determination.
   - Message Announce sent by every clock source carries the quality parameters clock-class and clock-accuracy required for clock source selection. Attribute LocalPriority is configured and is not transmitted over a packet.
   - If several-channel time sources have the same quality parameters, such as clock-class, clock-accuracy, and priority2, attribute LocalPriority can be used to compare the time source quality.
4. Clock attribute priority2 is configurable. It is 128 by default for a T-BC device.

Other clock attributes: clockAccuracy and offsetScaledLogVariance.

- clockAccuracy: indicates the clock's precision.
- offsetScaledLogVariance: indicates the clock's stability. It is the quality of clock signals that are received by an NE. The NE obtains the value using Allan deviation.

 Copyright © Huawei Technologies Co., Ltd.

14.5.3 Configuring 1588v2 (Based on G.8275.1)

Prerequisites

Time synchronization can be performed only after the clock has been synchronized. For example, synchronous Ethernet clock is used to implement clock synchronization.

1588v2 (based on G.8275.1) cannot be interconnected with 1588v2 (based on IEEE).

Procedure

Step 1 Run the **ptp device-type t-bc** command to set the device type to T-BC.

By default, the BC mode is used. That is, the device complies with the IEEE 1588v2 standard by default and implements 1588v2 based on the default profile. In this case, you need to manually switch the mode to T-BC to implement 1588v2 based on the ITUT G.8275.1 standard.

Step 2 Configure the global precision time protocol (PTP) function.

1. Run the **ptp t-bc** command to enter T-BC mode.
2. Run the **ptp enable** command to enable the PTP function on a device.
3. **Optional:** Run the **ptp domain** command to specify a time domain.

   After the configuration, a time domain has a clock source. All devices in the time domain use the clock source.
4. **Optional:** Run the **ptp virtual-clock-id** command to configure the virtual clock ID of a device.

   The clock ID uniquely identifies a clock node in a time domain.

   If this parameter is not configured, its value consists of the organizational unique identifier (OUI)+product ID+last four bytes of the MAC address. The OUI is 00-25-9e, and the product ID is 40. For example, if the MAC address of the system is 00e0-fc00-a025, the corresponding default clock ID is 00259e405898a025.

Step 3 Enable PTP on a port to transmit PTP packets.

1. Run the **ptp port frameid/slotid/portid enable** command to enable PTP on an Ethernet port.

   Enable 1588v2 on the ports that are required to transmit 1588v2 packets, for example, the uplink port on an OLT, user-side port on an MDU and GE port connecting the OLT and MDU.

   You are not required to enable the function on the GPON port connecting the OLT and MDU.
2. Run the **ptp port frameid/slotid/portid notslave disable** command to set the port to slave state. Then, you can trace the time source of the upper-layer device through this port.

   When the GE/10GE upstream port serves as the port for tracking the time source of the upper-layer device, it needs to be set to the slave state.

 Copyright © Huawei Technologies Co., Ltd.

However, this state does not need to be set for the GE port on the user side of the OLT and MDU and for the PON port between the OLT and MDU.

Step 4 Configure the T-BC time source for time synchronization on the device.

1. Run the **ptp source** command to configure the T-BC time source.

   The time source can be the T-BC time or BITS time recovered from the network-side GE/10GE/PON line. If the network does not support T-BC time synchronization, you can use the BITS time.

2. **Optional:** When the BITS time is used, run the **ptp bits** command to configure the attributes of a BITS clock source, such as the precision and class of the clock source, and the priority of the clock signal.

3. **Optional:** Configure the local priority of the time source.

   Time source selection supports best master clock algorithm (BMCA) for automatic selection. You can run the **ptp port frameid/slotid/portid local-priority** command to configure the local priority of the time clock source. If several-channel time sources have the same time attributes, you can configure the local priority to compare time source quality. The smaller the **local-priority**, the higher the priority.

Step 5 **Optional:** Configure PTP packet parameters.

Configure parameters carried in the PTP packet, including the timestamp mode and MAC mode.

1. Run the **ptp port frameid/slotid/portid clock-step { one-step | two-step }** command to configure the step mode for PTP ports.

   By default, 1588v2 packets carry timestamps in one-step mode. The 1588v2 ports identify Follow_Up packets in two-step mode in the Rx direction for communicating with other products.

2. **Optional:** Configure the encapsulation mode for PTP packets.

   Run the **ptp port frameid/slotid/portid mac-egress destination-mac destination-mac-value** command to configure the MAC encapsulation mode for PTP packets to be forwarded by the port. The packets are forwarded in Layer 2 mode.

   There are two types of destination multicast MAC addresses: unforwardable multicast MAC address and forwardable multicast MAC address.

Step 6 Configure asymmetric compensation parameters for optical fibers.

Run the **ptp port frameid/slotid/portid line-warp linewarptype linewarpdir linewarpvalue** command to configure asymmetric compensation parameters for optical fibers.

Asymmetric optical fibers between two devices result in time difference for data transmission and reception. During site deployment, measure and calculate the time differences in the two directions and compensate the time on the devices by configuring cable transmission deviations.

Step 7 **Optional:** Configure the asymmetric measurement function on a ring network.

Run the **ptp asymmetry-measure enable** command to configure the asymmetric measurement function on a ring network.

 Copyright © Huawei Technologies Co., Ltd.

This function utilizes the BMCA algorithm in G.8275.1. When fiber disconnection occurs on the ring network, services are automatically switched to the standby link, and the time synchronization precision of the nodes on the ring network stays in the usable range. When the faulty fiber is re-connected, the devices first automatically calculate and report asymmetry on the new fiber link and then compensate for asymmetry on the NMS. After that, services are switched back to the active link.

----End

Example

An MDU connects to a 4G LTM eNodeB using a GE port. The MDU connects to the OLT using the GPON uplink port and then to the radio network controller (RNC) over the upper-layer network to carry 4G services over the access network. The 4G LTE eNodeB requires high precision time synchronization and the 1588v2 time is deployed on the network.

The parameters planned on the OLT are as follows:

- Time domain: 24
- Clock or time source input: GE uplink port 0/8/0
- Clock or time source output: GPON service port 0/3/1
- Packet encapsulation mode: MAC (default mode) (packets are forwarded in Layer 2 multicast mode)

The parameters planned on the MDU are as follows:

- Time domain: 24
- Clock or time source input: GPON uplink port 0/0/0
- Clock or time source output: GE service port 0/1/1
- Packet encapsulation mode: MAC (default mode) (packets are forwarded in Layer 2 multicast mode)

Configurations on the OLT

```
huawei(config)#ptp device-type t-bc
huawei(config)#ptp t-bc
huawei(config-t-bc)#ptp enable
huawei(config-t-bc)#ptp port 0/8/0 enable
huawei(config-t-bc)#ptp port 0/8/0 notslave disable
huawei(config-t-bc)#ptp source 0 0/8/0
```

Configurations on the MDU

```
huawei(config)#ptp device-type t-bc
huawei(config)#ptp t-bc
huawei(config-t-bc)#ptp enable
huawei(config-t-bc)#ptp port 0/1/1 enable
huawei(config-t-bc)#ptp source 0 0/0/0
```

14.5.4 1588v2 (Based on G.8275.1) Reference Standards and Protocols

 Copyright © Huawei Technologies Co., Ltd.

Document	Description
G.8275.1	Precision time protocol telecom profile for phase/time synchronization with full timing support from the network.
G.8273.2	Timing characteristics of telecom boundary clocks and telecom time slave clocks. **NOTE** For details, see *Feature Specifications and Limitations*.
ITU-T G.813	Timing requirements of SDH equipment slave clocks (SEC)
ITU-T G.823	The control of jitter and wander within digital networks which are based on the 2048 kbit/s hierarchy
ITU-T G.8260	Definitions and terminology for synchronization in packet networks
ITU-T G.8261	Timing and Synchronization aspects in Packet Networks
ITU-T G.8262	Timing characteristics of Synchronous Ethernet Equipment slave clock (EEC)
G.987.3	XG-PON TC Layer Specification
G.988	ONU management and control interface specification (OMCI)

14.6 1588 ACR

1588 Adaptive Clock Recover (ACR) enables a master device that supports IEEE 1588v2 to encapsulate the local system clock information into 1588v2 packets and transparently send them to peer slave devices through a third-party network. The slave devices obtain the timestamp from the 1588v2 packets and restore the clock to implement frequency synchronization between devices at both ends of a packet switched network (PSN).

14.6.1 Why Is 1588 ACR Required

Benefits

On the current 3G and 4G LTE networks, the frequency recovery modes of devices at the convergence layer are synchronous Ethernet and 1588v2. In addition, the frequency recovery is performed hop by hop. Therefore, the hardware of the entire network must support synchronous Ethernet or 1588v2, which has high requirements on the hardware of the network.

However, in the 2G mobile bearer network, enterprise private line, and narrowband service network reconstruction scenarios, the MAN cannot be

 Copyright © Huawei Technologies Co., Ltd.

reconstructed to provide hop-by-hop clock synchronization. Therefore, it is difficult to introduce accurate clock synchronization.

How to achieve high clock synchronization accuracy with low network deployment costs?

The IEEE proposes the 1588 ACR solution, that is, the adaptive clock recovery solution. In this solution, frequency information is carried in 1588v2 packets and transparently transmitted over an intermediate network. Devices at both ends recover frequency information using 1588v2 packets to implement frequency synchronization.

Function

1588 ACR synchronizes the frequencies of the devices at the two ends of a packet switched network (PSN). Specifically, a master device supporting 1588v2 encapsulates the local system clock data into 1588v2 packets. The PSN network transparently transmits the 1588v2 packets to a slave device. The slave device obtains timestamps from the received 1588v2 packets and recovers the clock data of the master device. The PSN network does not need to support clock synchronization, or the clock of the PSN network can be a third-party clock.

1588 ACR implements frequency synchronization but not time synchronization.

14.6.2 1588 ACR Basic Concepts

The basic concepts involved in the 1588 adaptive clock recovery (ACR) feature are the same as those of 1588v2. For details, see **14.4.2 1588v2 Basic Concepts** of the 1588v2 feature.

14.6.2.1 1588 ACR Messages

ITU-T G.8265.1 defines seven types of 1588 adaptive clock recovery (ACR) messages. The **Table 14-9** describes each type of 1588 ACR messages.

Table 14-9 1588 ACR messages

Type	Classification	Message Transmit Rate (pps)	Function
Sync	Event message	1/16-128	Sync, Delay_Req, Follow_Up, and Delay_Resp messages carry timestamps for recovering clock frequencies.
Delay_Req			
Follow_Up	General message	1/16-128	
Delay_Resp			

 Copyright © Huawei Technologies Co., Ltd.

Type	Classification	Message Transmit Rate (pps)	Function
Announce		1/16-8 Default: 1/2	Announce messages carry best master clock adaptive (BMCA) information, such as priorities and quality levels (QLs). The Announce messages are used by slave devices to select a clock source and by master devices to implement clock protection switchovers.
Signaling		None	Signaling messages are used for unicast negotiations. A slave device sends a master device a signaling message to request Sync or Announce messages. Then the master device sends a signaling message to respond to the slave device. The master and slave devices exchange Sync or Announce messages only after the negotiation between them is successful.
Management		N/A	Management messages are used for device management. ITU-T G.8265.1 has not defined how to use the management messages.

ITU-T G.8265.1 defines only transmit rate ranges for 1588v2 messages. Select a message transmit rate based on device hardware capability (crystal oscillation) and clock performance requirements (usage scenario and customer's requirements). A higher message transmit rate results in better clock recovery performance. The message transmit rate is configured on a slave device. After the configuration, the slave device sends the message transmit rate to the master device through a signaling request. Then the master device sends Sync or Announce messages to the slave device at the message transmit rate.

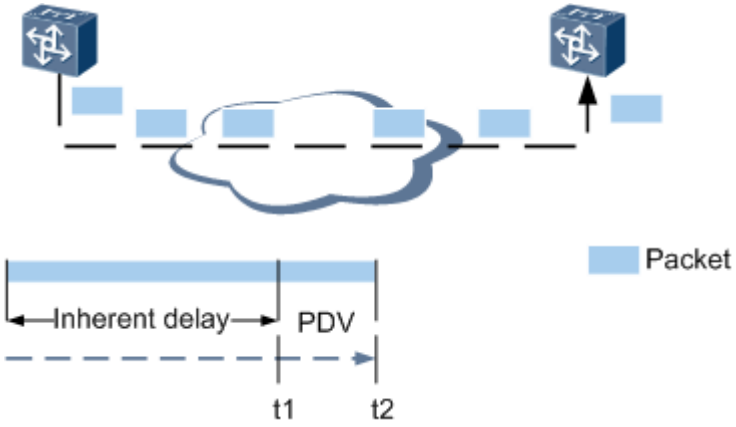
14.6.2.2 PDV

1588 adaptive clock recovery (ACR) is related to packet delay variation (PDV), which indicates the delay variation of the packet transmitted over a network. As shown in **Figure 14-22**, the packet forwarding delay is calculated using the following formula: Delay = Inherent delay + PDV.

 Copyright © Huawei Technologies Co., Ltd.

- Inherent delay: the minimum delay when packets pass through a network. The inherent delay has a fixed value, which is the sum of the physical link (for example, optical fiber) delay and the fixed circuit delay inside a device.
- PDV: Because of queue scheduling by priority and clock domain translation, a device has to store packets in buffer registers, such as first in first out (FIFO) and random access memory (RAM), for a period of time (for example, several milliseconds) before sending them. As a result, the packets arrive at a destination after a delay, at any time from t1 to t2 in **Figure 14-22**. The delay is called PDV.

Figure 14-22 PDV diagram



14.6.2.3 1588 ACR Modes

Device Model

For the concept of device model, see **14.4.2.1 Device Model**.

According to ITU-T G.8265.1, a slave device supporting 1588 adaptive clock recovery (ACR) supports only the slave-only ordinary clock (SOOC) model.

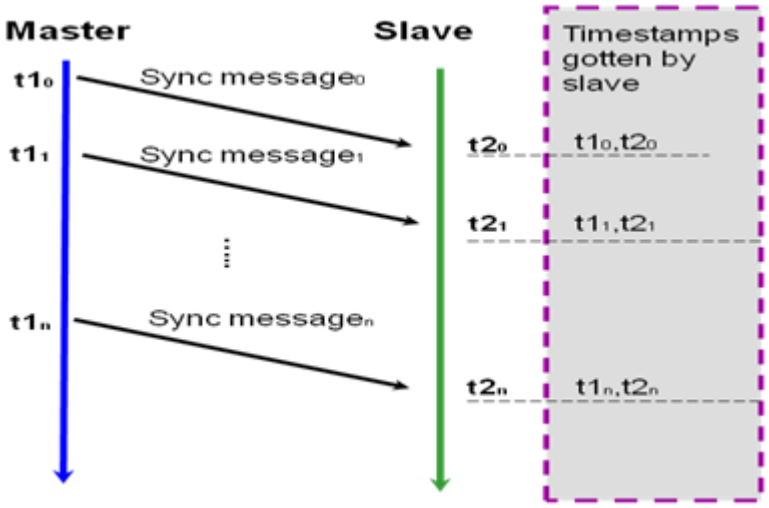
The 1588 ACR feature applies in end-to-end (E2E) deployment scenarios. A slave device needs to synchronize its clock frequency only with that of a master device. Therefore, the slave device does not need to support the boundary clock (BC) model.

One-way and Two-way

In one-way mode, the master device sends Sync messages to the slave device and the slave device does not send Delay_Req messages to the master device. The slave device recovers the frequency according to the timestamp contained in the Sync messages. For details, see **Figure 14-23**.

Copyright © Huawei Technologies Co., Ltd.

Figure 14-23 One-way mode



The master periodically sends Sync messages to the slave. If the slave frequency is synchronized to the master frequency, then the accumulative time errors within the same time periods are the same, as long as the path delay changes are neglected.

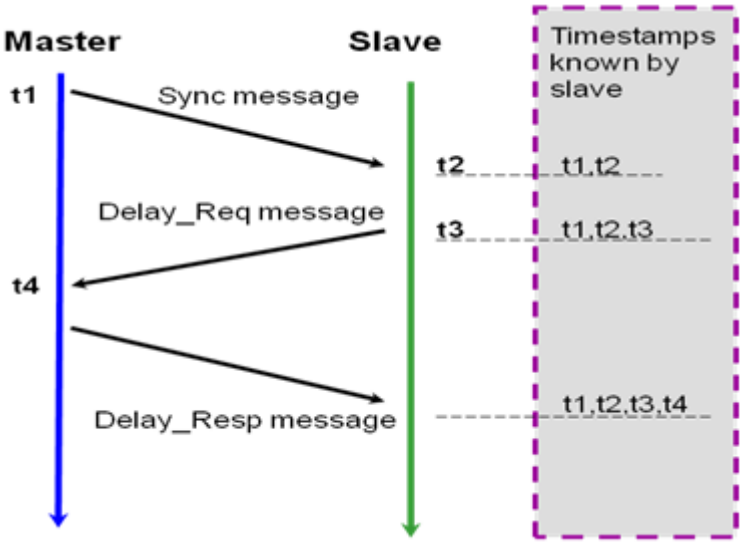
In other words, $t2_1$-$t2_0$=$t1_1$-$t1_0$, $t2_2$-$t2_1$=$t1_2$-$t1_1$, $t2_3$-$t2_2$=$t1_3$-$t1_2$..., $t2_n$-$t2_0$=$t1_n$-$t1_0$.

If $t2_n$-$t2_0$ is greater than $t1_n$-$t1_0$, then the slave frequency is higher than the master frequency, which means the slave frequency must be decreased. Reversely, the slave frequency must be increased.

One-way is the basic principle of 1588 ACR frequency synchronization, and can meet the synchronization requirements of 1588 ACR.

In two-way mode, packets are transmitted bidirectionally between the master and slave. The master sends Sync packets to the slave, and the slave sends Delay_Req packets to the master. The master and slave exchange 1588v2 packets to implement frequency synchronization. For details, see **Figure 14-24**.

Figure 14-24 Two-way



 Copyright © Huawei Technologies Co., Ltd.

1. The master sends a Sync message at t1 and carries the t1 timestamp in the Sync message.
2. The slave receives the Sync message at t2, locally generates the t2 timestamp, and extracts the t1 timestamp from the Sync message.
3. The slave sends a Delay_Req message at t3 and locally generates the t3 timestamp.
4. The master receives the Delay_Req message at t4, locally generates the t4 timestamp, and sends the Delay_Req message with the t4 timestamp back to the slave.
5. The slave extracts the t4 timestamp from the Delay_Resp message after receiving it.

"Delayms" is the path delay in the direction from the master to the slave, "Delaysm" is the path delay in the direction from the slave to the master, and "Offset" is the time offset between the slave and master. All of them are variables.

t2-t1=Delayms+Offset

t4-t3=Delaysm-Offset

When Delayms = Delaysm, that is, when the transmit and receive links between the master and slave are symmetric, the following formula is satisfied:

Offset=[(t2-t1)-(t4-t3)]/2

The slave can calculate the time offset between itself and the master based on the t1, t2, t3, and t4 timestamps and then corrects its own time to get synchronized with the master.

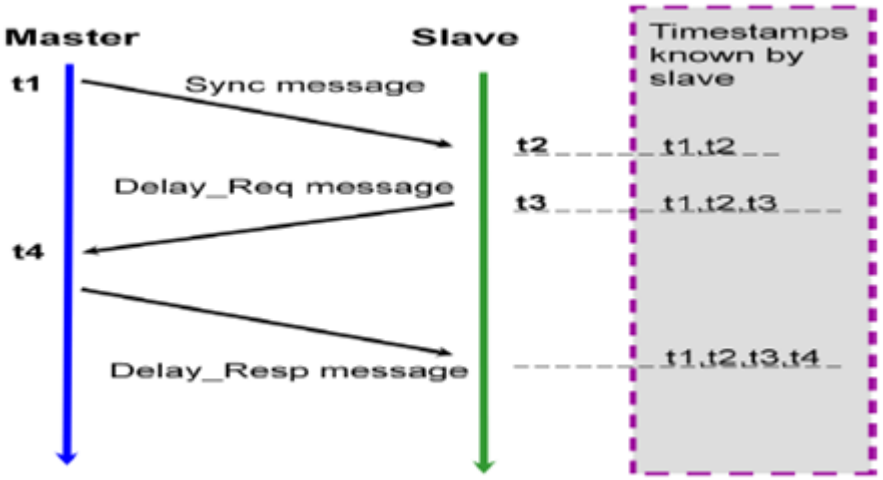
Both the one-way and two-way modes implement frequency synchronization and the one-way mode uses fewer bandwidths. According to ITU-T G.8265.1, the master must support both one-way and two-way modes. A slave is required to support either the one-way or two-way mode. The master can be connected to slaves that support the one-way or two-way mode.

One-step and Two-step

In One-step mode, there is no follow-up packet, and the Sync packet directly carries the timestamp t1, as shown in **Figure 14-25**.

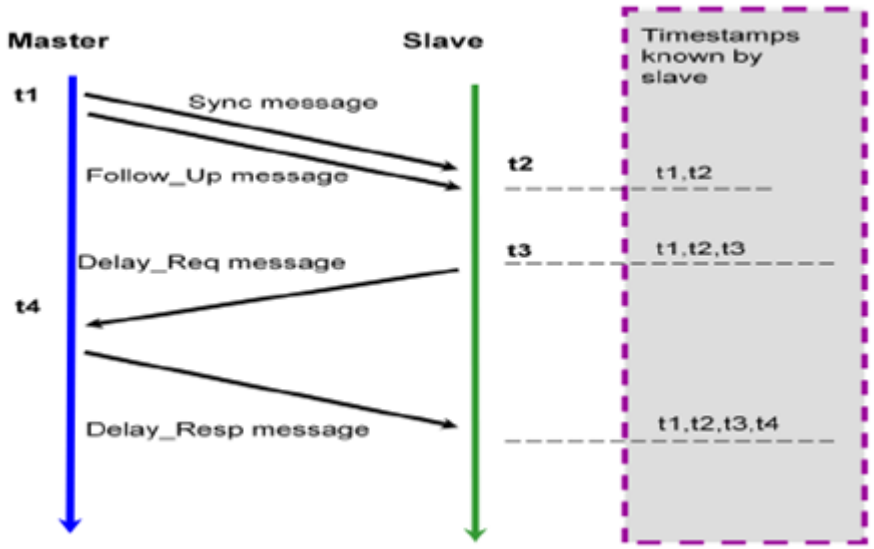
 Copyright © Huawei Technologies Co., Ltd.

Figure 14-25 One-step mode



In two-step mode, each time a Sync packet is sent, a follow_up packet is generated to carry the timestamp t1 generated when the Sync packet is sent, as shown in **Figure 14-26**.

Figure 14-26 Two-step mode



The two-step mode is defined by IEEE 1588v2. In this mode, a Sync message corresponds to a Follow_Up message. When a Sync message passes the timestamp point, the master device hardware generates timestamp t1. The master device does not insert timestamp t1 into the Sync message, but inserts it into the Follow_Up message. The slave device extracts timestamp t1 from the received Follow_Up message. This reduces real-time requirements for the master and slave devices.

According to ITU-T G.8265.1, a 1588 ACR master device supports either or both of the one-step and two-step modes, and a 1588 ACR slave device automatically

 Copyright © Huawei Technologies Co., Ltd.

adapts to the mode carried in a Sync message sent from the master device. If the slave device connects to two master devices that use different 1588 ACR modes, the slave device must support the two modes.

Packet Encapsulation

According to ITU-T G.8265.1, 1588 ACR messages are encapsulated in UDP over IPv4 mode. This meets the requirements of appendix D *Transport of PTP over User Datagram Protocol over Internet Protocol Version 4* in the IEEE 1588v2 protocol.

1588 ACR messages are unicast UDP over IPv4 packets.

14.6.2.4 1588 ACR Clock Source Selection Algorithm

BMCA Algorithm

Based on the best master clock (BMC) algorithm of IEEE 1588v2, ITU-T G.8265.1 defines the best master clock adaptive (BMCA) algorithm for 1588 adaptive clock recovery (ACR). BMCA is simpler than BMC.

Table 14-10 lists BMCA parameters involved in 1588 ACR clock source selection.

Table 14-10 BMCA parameters involved in 1588 ACR clock source selection

Parameter	Description
Packet timing signal fail (PTSF)	Indicates that a Sync message is lost, an Announce message is lost, or a clock source is unavailable. The possible cause of an unavailable clock source is as follows: The packet delay variation (PDV) or the clock jitter of the master device is so large that the slave device cannot tolerate.
Quality level (QL)	Each master device carries its QL through an Announce message. The slave device preferentially traces the clock of the master device with the highest QL.
Priority	A slave device configures a local priority for each master device. If multiple master devices have the same QL, the slave device traces the clock of the master device with the highest priority.

If a slave device connects to multiple master devices and links are set up between the slave device and the master devices through unicast negotiations, the slave device receives the Announce and Sync messages sent from each master device. Then the slave device determines the best clock source as follows:

1. The slave device checks the PTSF of the clock of each master device. If PTSF occurs on a master device, the slave device does not trace the clock of the master device.

 Copyright © Huawei Technologies Co., Ltd.

2. The slave device compares the QL of each master device and traces the clock of the master device with the highest QL.
3. If multiple master devices have the same QL, the slave device traces the clock of the master device with the highest priority.

Relationship Between 1588 ACR Clock Synchronization and Physical-Layer Clock Synchronization

The relationships between the 1588v2 clock synchronization and the physical-layer clock synchronization, including SDH synchronization and synchronous Ethernet, are as follows:

- The clock source of a 1588 ACR master device may be from an SDH or a synchronous Ethernet network.

  In this scenario, the master device must convert the synchronization status message (SSM) QL of the SDH or synchronous Ethernet network to the 1588 ACR clock class. In this manner, the clock sources of the master devices can be switched over for protection. In addition, a master device sends its clock QL to the slave device through Announce messages.
- The clock output from a 1588 ACR slave device can be used as the clock input to an SDH or synchronous Ethernet network.

  In this scenario, the 1588 ACR slave device recovers the clock frequency from 1588v2 messages and synchronizes the clock frequency with that of the master device. In addition, the slave device provides the recovered clock frequency to the downstream network or base station through the SDH or synchronous Ethernet network. In this case, the slave device must convert the 1588 ACR clock class to the SSM QL of the SDH and synchronous Ethernet network. In this manner, the clock sources of the slave devices can be switched over for protection.

ITU-T G.8265.1 defines the mapping between the 1588 ACR clock class and the SSM QL of the SDH or synchronous Ethernet network, as shown in the **Table 14-11**.

Table 14-11 Mapping between the 1588v2 clock class and the SSM QL

SSM QL	ITU-T G.781 Option Synchronization layer functions	1588v2 Clock Class
0010	QL-PRC	84
0100	QL-SSU-A	90
1000	QL-SSU-B	96
1011	QL-SEC	104
1111	QL-DNU	110

14.6.3 1588 ACR Principles

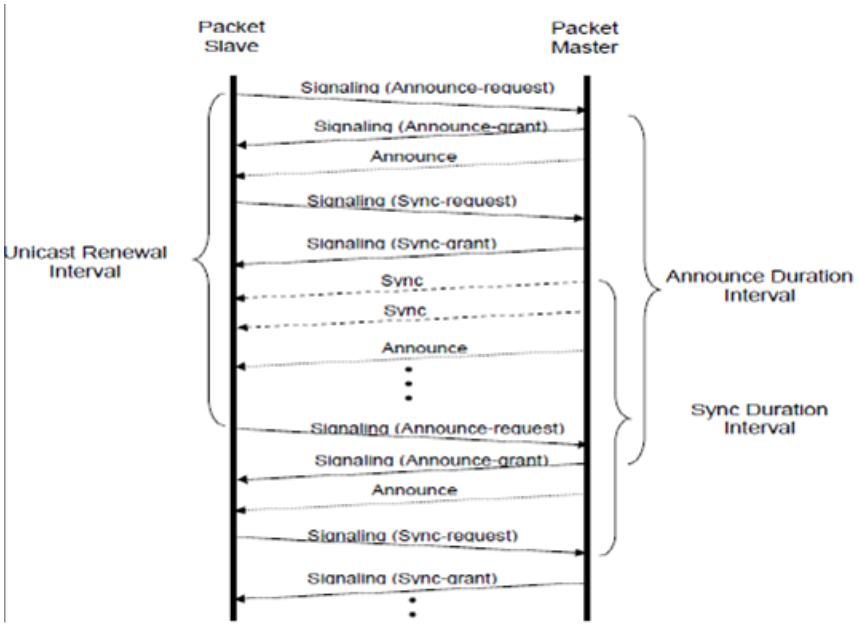
 Copyright © Huawei Technologies Co., Ltd.

14.6.3.1 1588 ACR Unicast Negotiation

ITU-T G.8265.1 defines that a slave device communicates with a master device through a unicast negotiation. Specifically, the slave device initiates a request to the master device. The master device sends 1588v2 packets to the slave device only after authenticating the slave device.

A unicast negotiation is performed through signaling messages that carry request data. The request data includes the requested message type, message transmit rate, and duration. The duration defines the valid period of a request. The master device sends the messages requested by the slave device only within the duration. After the duration elapses, the master device stops sending the messages to the slave device. The master device will start sending messages again only after the slave device initiates a new request. **Figure 14-27** shows the process of a 1588 adaptive clock request (ACR) unicast negotiation.

Figure 14-27 Process of a 1588 ACR unicast negotiation



1. The slave device sends the master device a signaling message to request Announce messages. The signaling message contains the type, message transmit rate, and duration of the requested Announce messages.
2. The master device sends a signaling message to respond to the slave device, notifying the slave device that the request is permitted or rejected.
   - If the request is permitted, the master device sends Announce messages to the slave device at the message transmit rate requested by the slave device.
   - If the request is rejected, the master device does not send Announce messages to the slave device.

 Copyright © Huawei Technologies Co., Ltd.

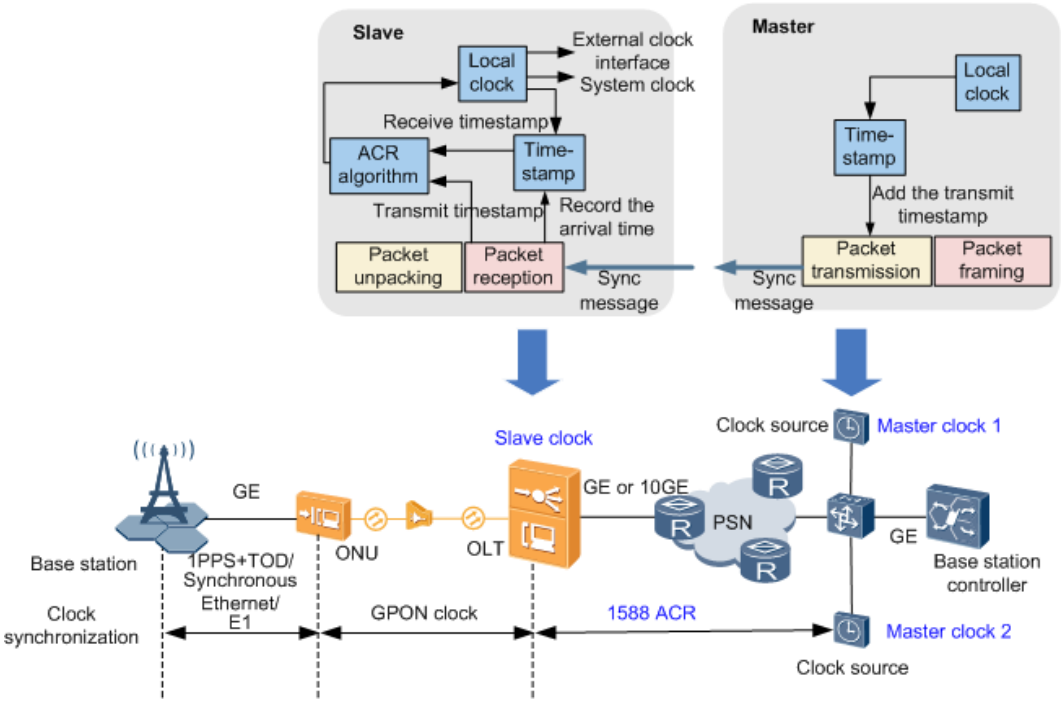
3. The slave device sends the master device a signaling message to request Sync messages. The signaling message contains the type, message transmit rate, and duration of the requested Sync messages.
4. The master device sends a signaling message to respond to the slave device, notifying the slave device that the request is permitted or rejected.
   - If the request is permitted, the master device sends Sync messages to the slave device at the message transmit rate requested by the slave device.
   - If the request is rejected, the master device does not send Sync messages to the slave device.

ITU-T G.8265.1 defines that the duration of each message can be configured on the slave device. The duration ranges from 60s to 1000s, with 300s by default. Before duration elapses, the slave device determines whether to initiate a new request. To ensure message continuity, the slave device initiates a new request after the duration elapses and before the master device stops sending messages.

14.6.3.2 1588 ACR Principles

1588 adaptive clock recovery (ACR) uses the ACR algorithm to recover the clock frequency of the transmit end from the transmit and receive timestamps carried in 1588v2 packets.

Figure 14-28 1588 ACR principles



The 1588 ACR process is as follows:

1. The slave device uses the 1588 ACR clock source selection algorithm to determine the best master clock from master devices 1 and 2 for tracing.

 Copyright © Huawei Technologies Co., Ltd.

2. The master device providing the best master clock adds a transmit timestamp that contains the local system time to a 1588v2 packet.
3. The packet switched network (PSN) transparently transmits the 1588v2 packet to the slave device.
4. The slave device extracts the transmit timestamp from the received 1588v2 packet and adds the receive timestamp to the 1588v2 packet.
5. The slave device uses the ACR algorithm to process the transmit and receive timestamps contained in the 1588v2 packets that are received within the specified duration. Then the slave device recovers the clock frequency of the master device.

14.6.4 1588 ACR Deployment Requirements

When planning and deploying 1588 adaptive clock recovery (ACR), minimize the packet delay variation (PDV) for a stable running environment.

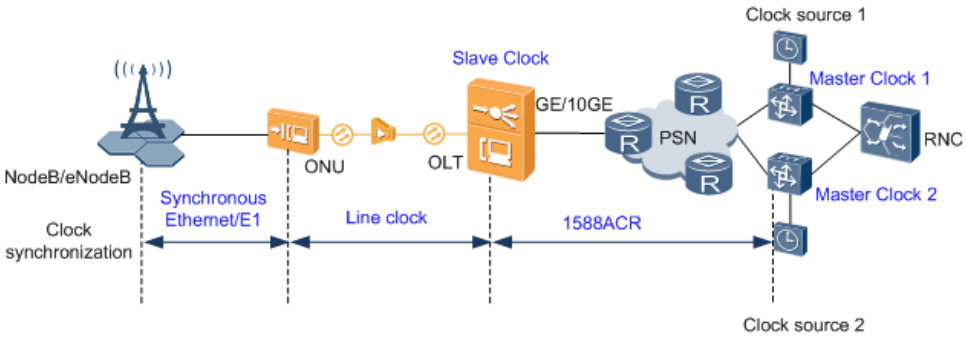
- Clock source protection: Deploy at least two master devices to implement clock source protection switchover.
- Locations of master and slave devices: Deploy the master and slave devices on the edge nodes connected to the intermediate network. This reduces the PDV caused by the intermediate network.
- Scale of an intermediate network: Adjust the routes of 1588 ACR messages and the locations of master and slave devices to minimize the number of forwarding hops for the 1588 ACR messages.
- Type of an intermediate network: Test and collect the PDV data on the live network to check whether 1588 ACR can be deployed on the network.
- Network traffic: Ensure that the long-term average traffic of the network is less than 80% of the maximum traffic that can be afforded by the network. Short-term network congestion and service interruptions are allowed according to G.8261.
- Quality of service (QoS) priority: Configure a high priority for 1588 ACR messages to reduce the PDV caused by packet forwarding.
- Packet loss ratio: Ensure that the packet loss ratio is not greater than 0.5%. This prevents timing signal failure.
- Message transmit rate:
  - A higher transmit rate of Sync messages results in better clock recovery performance.
  - A higher transmit rate of Announce messages results in faster fault detection and protection switchover.
  - If there are many slave devices, a higher message transmit rate results in a higher bandwidth. In this case, the one-step mode is recommended. If the bandwidth is still high after the one-step mode is used, decrease the message transmit rate.

14.6.5 1588 ACR Networking

Figure 14-29 shows the 1588 adaptive clock recovery (ACR) networking.

 Copyright © Huawei Technologies Co., Ltd.

Figure 14-29 1588 ACR networking



- Clock sources 1 and 2 inject clock signals to the primary and secondary master devices, respectively. The clock signals are transmitted through 1588v2 or synchronous Ethernet. The convergence layer is a third-party packet switched network (PSN) or an intermediate network that does not support 1588v2 or synchronous Ethernet. 1588 ACR is deployed on the convergence layer to ensure clock precision.
- The optical line terminal (OLT), a slave device, supports the negotiation with two master devices. According to the 1588 ACR clock source selection algorithm, the OLT traces the clock of the master device that provides better clock signals.
- The master devices convert 1588v2 or synchronous Ethernet clock packets to timestamps and encapsulate the timestamps into 1588 ACR messages. The convergence layer transparently transmits the 1588 ACR messages to the OLT. The OLT recovers the clock frequency from the received 1588 ACR messages and sends the clock frequency to the optical network unit (ONU) through line clock synchronization.
- The ONU sends the clock frequency to the base station through synchronous Ethernet, or E1 line clock synchronization.

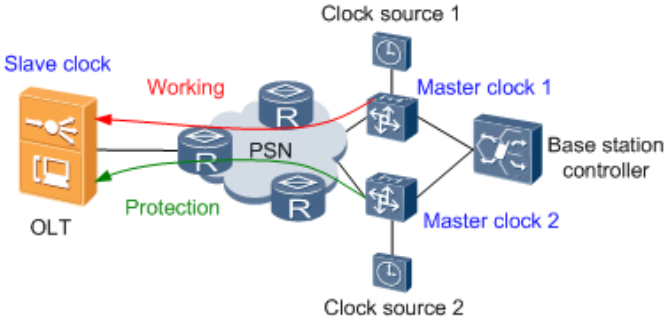
The OLT is located at the end of a mobile bearer network or an access network. If 1588 ACR is enabled, the OLT can be used only as a slave device.

1588 ACR Protection Networking

If a clock source is faulty, services are interrupted. Therefore, the clock source must be protected. For the 1588 ACR feature, deploy a secondary master device to provide a secondary clock source. **Figure 14-30** shows the 1588 ACR protection networking.

 Copyright © Huawei Technologies Co., Ltd.

Figure 14-30 1588 ACR protection networking



In the preceding networking, if the primary master device is faulty, the slave device automatically traces the clock of the secondary master device. After the primary master device recovers, the slave device automatically traces the clock of the primary master device.

14.6.6 Configuring 1588 ACR

This section describes how to configure 1588 adaptive clock recovery (ACR) to implement the end to end (E2E) frequency synchronization across a third-party network.

Prerequisites

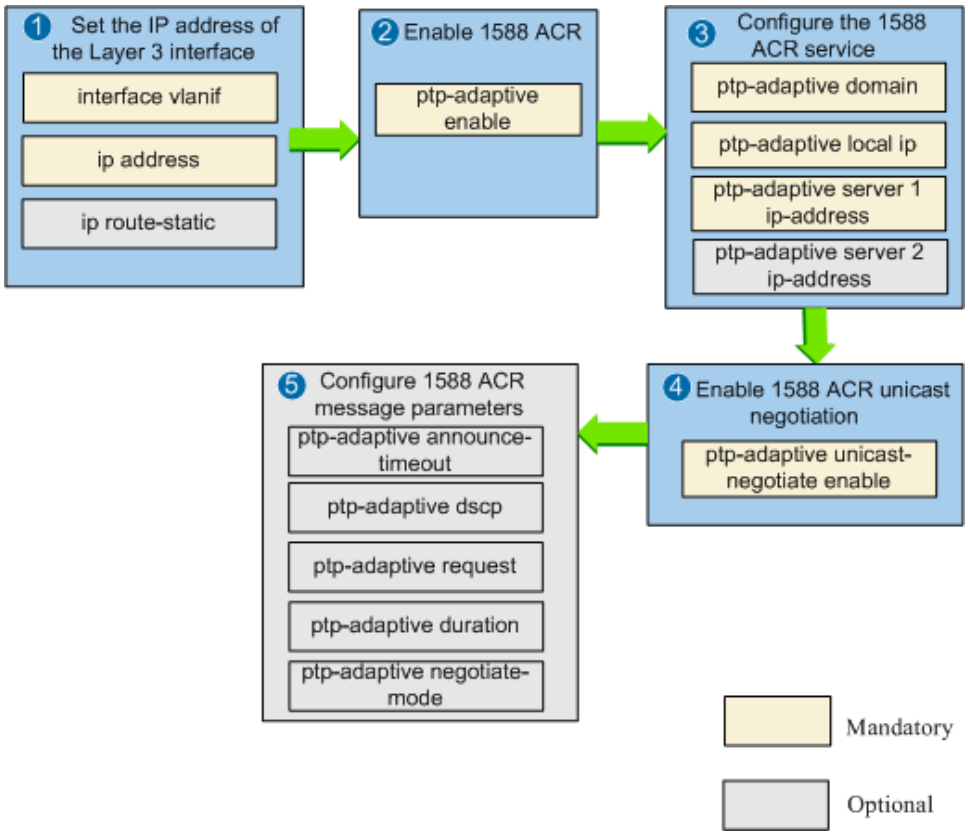
- The optical line terminal (OLT) has been installed and deployed.
- The master device is functional.
- The route on the third-party network is reachable.

Configuration Process

Figure 14-31 shows the 1588 ACR configuration process.

 Copyright © Huawei Technologies Co., Ltd.

Figure 14-31 1588 ACR configuration process



Procedure

Step 1 Set the IP address of the Layer 3 interface.

1. Run the **interface vlanif** command to create a VLANIF interface and enter the VLANIF mode.
2. Run the **ip address** command to set the IP address of the VLAN interface.

   NOTE

   Ensure that the IP address of the VLAN interface and that of the master device port are in the same network segment.
3. **Optional:** Run the **ip route-static** command to configure a route from the local IP address of the master device to the IP address of the VLAN Layer 3 interface on the slave device.

   Perform this step if the local IP address of the master device port and that of the VLAN Layer 3 interface on the slave device are in different network segments.

Step 2 Run the **ptp-adaptive enable** command to enable 1588 ACR.

Step 3 Configure the 1588 ACR service.

1. Run the **ptp-adaptive domain** command to configure the time domain for the 1588 ACR devices.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

Clock synchronization can be performed on the slave device only if the master and slave devices are in the same time domain.

2. Run the **ptp-adaptive local-ip** command to configure the local IP address of the 1588 ACR slave device.

   1588 ACR messages are Layer 3 unicast IP packets. Therefore, a fixed IP address is required between the master and slave devices. This IP address is generally a loopback address or the IP address of an outbound port. In this step, the local IP address is the IP address of the VLAN interface set in **Step 1.2**.

3. Run the **ptp-adaptive server 1 ip-address** command to set the IP address of 1588 ACR master device server 1.

   NOTE

   In this step, the IP address is the local IP address of the master device.

4. **Optional:** Run the **ptp-adaptive server 2 ip-address** command to set the IP address of master device 2.

   Perform this step if a protection switchover is required for master devices.

   NOTE

   If a slave device connects to two master devices, the slave device traces the clock of the master device with a higher quality level and priority based on the 1588 ACR clock source selection algorithm. If the primary master device is faulty, the slave device automatically traces the clock of the secondary master device based on the 1588 ACR clock source selection algorithm.

Step 4 Run the **ptp-adaptive unicast-negotiate enable** command to enable 1588 ACR unicast negotiation.

Step 5 **Optional:** Configure 1588 ACR message parameters.

- Run the **ptp-adaptive announce-timeout** command to configure the maximum number of times that Announce messages are not received within a packet transmit period. The default number is 3.
- Run the **ptp-adaptive dscp** command to configure the priority for processing forwarded 1588 ACR messages. The default priority is 56.

  NOTE

  This parameter determines the 1588 ACR message forwarding priority. A higher priority results in a smaller network jitter for the 1588 ACR messages. Accordingly, the recovered clock performance is better. Otherwise, the recovered clock performance is not so good.

- Run the **ptp-adaptive request interval** command to configure the interval requested by the slave device at which the master device sends 1588 ACR Announce, Sync, or Delay_Req messages to the slave device.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

- Default interval for Announce messages: 11
- Default interval for Sync messages: 3
- Default interval for Delay_Req messages: 3

The minimum message transmit rate required by the frequency recovered through 1588 ACR varies according to network condition. A larger packet transmit rate results in better recovered frequency performance. However, this requires more system resources and bandwidths. Therefore, configure a proper value to balance the frequency performance, network resources, and bandwidths.

- Run the **ptp-adaptive duration** command to configure the interval at which a re-negotiation is performed between the master and slave devices. The default time is 300s.

  NOTE

  The duration for each type of packet is separately configured. Therefore, before a duration elapses, the slave device determines whether to initiate a new request based on site requirements. To ensure message continuity, the slave device initiates a new request after the duration elapses and before the master device stops sending messages.

- Run the **ptp-adaptive negotiate-mode** command to configure the 1588 ACR negotiation mode to one-way or two-way. The default mode is one-way.

  NOTE

  The frequency recovered in two-way mode is better than that recovered in one-way mode, but this mode requires more bandwidths.

----End

Result

After the **display ptp-adaptive all** command is executed, **Ptp adaptive state** is **Enable**, **Current state** is **Slave**, and **Nego-state** is **Success**.

If an error occurs in clock tracing, rectify the fault based on the reported clock alarm or event.

Example

The following configurations are used as an example to configure 1588 ACR: An OLT connects to a third-party network through GE upstream transmission and then to a radio network controller (RNC). This implements wireless service carried over a base station network. The third-party network does not support 1588v2. Therefore, 1588 ACR is deployed on the network.

Data plan on the OLT:

- Time domain: 1
- VLAN ID: 2
- IP address of the VLAN interface: 10.10.11.2
- IP address of 1588 ACR master device server 1: 10.10.11.10

OLT configuration

```
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.10.11.2 24
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-if-vlanif2)#quit
huawei(config)#ptp-adaptive enable
huawei(config)#ptp-adaptive domain 1
huawei(config)#ptp-adaptive local-ip 10.10.11.2
huawei(config)#ptp-adaptive server 1 ip-address 10.10.11.10
huawei(config)#ptp-adaptive unicast-negotiate enable
```

14.6.7 1588 ACR Maintenance and Diagnosis

1. If any alarms or events is reported, rectify faults using the alarm or event troubleshooting methods.
2. Run the **display ptp-adaptive all** command to query the 1588 ACR running status. Rectify faults, if any.
3. Run the **display ptp-adaptive server** command to query the statistics of the collected transmit and receive 1588 ACR messages. Rectify faults, if any.
4. Run the **display ptp-adaptive all** command to query 1588 ACR configurations. Ensure that 1588 ACR is enabled, the IP address of the 1588 ACR server is correct, and 1588 ACR attributes are correct.
5. Check clock hardware status. Ensure that the upstream interface board, physical ports and links, and the clock daughter board are functional. Specifically, no board or port alarm is reported. A clock daughter board fault can be queried through alarms or commands.
6. Run the **display ptp-adaptive all** command to check the status and configuration of the Layer 3 interface.
7. Run the **display ip routing-table** command to check whether the master and slave routes are reachable.

14.6.8 1588 ACR Standard and Protocol Compliance

Standard or Protocol	Description
IEEE 1588-2008	Precision Clock Synchronization Protocol for Networked Measurement and Control Systems
ITU-T G.8260	Definitions and terminology for synchronization in packet networks
ITU-T G.8261	Timing and Synchronization aspects in Packet Networks
ITU-T G.8262	Timing characteristics of Synchronous Ethernet Equipment slave clock (EEC)
ITU-T G.8265	Architecture and requirements for packet based frequency delivery
ITU-T G.8265.1	Precision time protocol telecom profile for frequency synchronization

 Copyright © Huawei Technologies Co., Ltd.

14.6.9 Remote 1588 Packet Capture

If a failure occurs during 1588v2 protocol interconnection, remote 1588v2 packet capture facilitates fault locating so that customers can quickly rectify a fault. This feature supports capture of 1588v2 (IEEE-based), 1588v2 (G.8275.1-based), and 1588 ACR packets.

Context

The remotely captured 1588 packets are first buffered on the control board, and uploaded to the file server as files , and then opened by using the tool.

Procedure

Step 1 Run the **capture 1588v2** command to enable 1588v2 packet capture.

If 1588 ACR packets are captured, run the **capture 1588acr** command to enable 1588 ACR packet capture.

Step 2 Run the **file-server auto-backup debug** command to configure the backup server.

Step 3 Run the **capture start** command to start packet capture.

Step 4 During packet capture, run the **capture stop** command to stop it. Then, the device uploads packets to the server.

Step 5 Once packet capture stops, run the **undo capture 1588v2** or **undo capture 1588acr** command to disable 1588 packet capture.

----End

Example

To capture 1588v2 (IEEE-based) packets, do as follows:

```
huawei(config)#diagnose
huawei(diagnose)%%capture 1588v2 0/2/0
huawei(diagnose)%%file-server auto-backup debug primary 192.168.1.20 sftp
huawei(diagnose)%%capture start -c 10000    //10000 indicates the number of captured packets.
huawei(diagnose)%%undo capture 1588v2 0/2/0
```

14.7 NTP

The Network Time Protocol (NTP) is used to synchronize the time between the distributed time server and the client.

NOTE

You are advised to enable NTP for accurate system time.

14.7.1 NTP Introduction

Definition

The Network Time Protocol (NTP) is an application layer protocol in the TCP/IP protocol suite. NTP is used to synchronize the time between the distributed time

 Copyright © Huawei Technologies Co., Ltd.

server and the client. The implementation of NTP is based on IP and UDP. NTP involves the Time Protocol and the ICMP Timestamp Message, with special design on accuracy and robustness.

Purpose

NTP defines the accurate time in an entire network.

Because the network topology is complicated, the clock synchronization among all the devices in the entire network becomes more critical. The objective of NTP is to synchronize the clocks of all the devices on a network which have clocks. This helps to keep time consistency among all the devices in the network. Therefore, the device can offer various applications based on the clock synchronization. For example, before analyzing logs collected from different devices for network management, you must ensure that the time on the logs is from synchronized clocks.

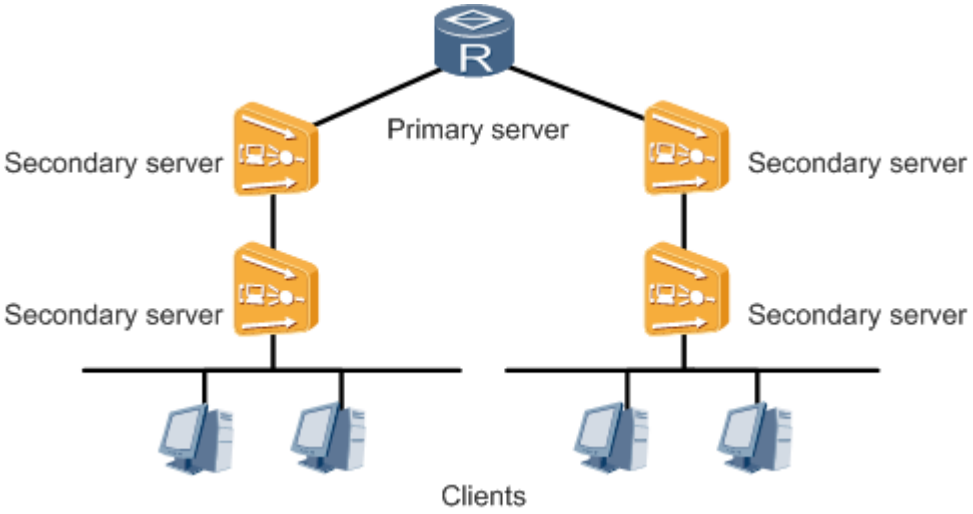
The MA5800 supports the NTP feature to guarantee that the clocks of all the devices in a network are consistent.

14.7.2 NTP Principle

14.7.2.1 NTP Network Architecture

As shown in the followed figure, the networking of NTP is composed of a primary time server, secondary time servers, clients, and interconnecting transmission paths.

Figure 14-32 NTP network architecture



- A primary time server directly synchronizes its clock with a primary reference source which is usually a radio clock or Global Positioning System (GPS).
- A secondary time server synchronizes its clock with the clock of the primary time server or other secondary time servers on the network, and transmits the time information to clients on the network through NTP.

 Copyright © Huawei Technologies Co., Ltd.

Under normal circumstances, primary and secondary time servers in a synchronization subnet assume a hierarchical-master-slave structure, with the primary server at the top and the secondary servers at successive stratums near leaf nodes. Secondary time servers at stratums further from the master time server will get less accurate clock information.

14.7.2.2 NTP Operating Mode

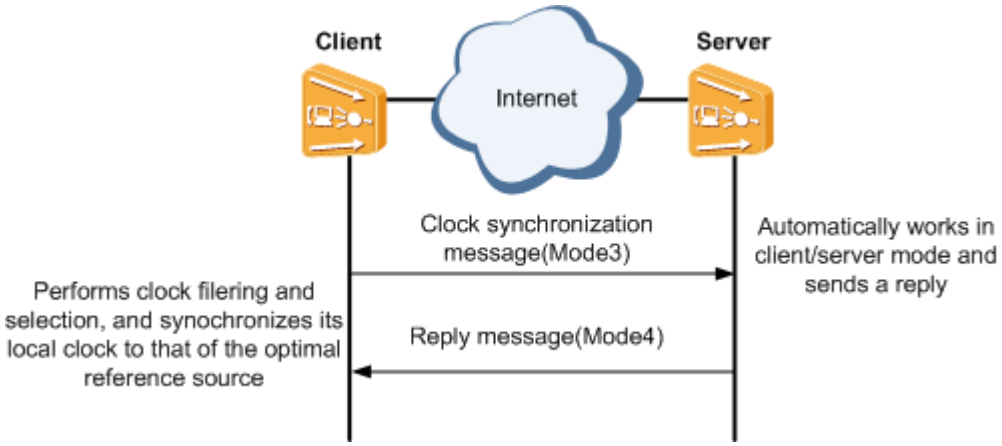
In actual application, you need to select a proper NTP operating mode based on the network deployment to meet various clock synchronization requirements. The operating modes of NTP are classified into Unicast Client/Server Mode, peer mode, broadcast mode, multicast mode and manycast mode.

Unicast Client/Server Mode

- The host that functions as a client sends packets to the server periodically. The value of the Mode field in a packet is set to 3. This indicates that the packet is sent by a client, without considering whether the server is reachable and which stratum the server is on. Usually, the host operating in client mode is a workstation on a specified network, which synchronizes its clocks with the clock on the server but does not alter the clock of the server.
- The host that functions as a server receives the packets from the client and sends response packets. The value of the Mode field in a response packet is set to 4. This indicates that the packet is sent by a server. Usually, the host operating in server mode is a time server on a network, which provides synchronization information for the clients but does not alter its own clock.

During and after the restart, the host operating in client mode periodically sends NTP request messages to the host operating in server mode. After receiving the NTP request message, the server swaps the position of destination IP address and source IP address, and the source port number and destination port number, fills in the necessary information, and sends the message to the client. The server does not need to retain state information when the client sends the request message. The client freely adjusts the interval for sending NTP request messages according to the local conditions.

Figure 14-33 Unicast Client/Server Mode



 Copyright © Huawei Technologies Co., Ltd.

Peer Mode

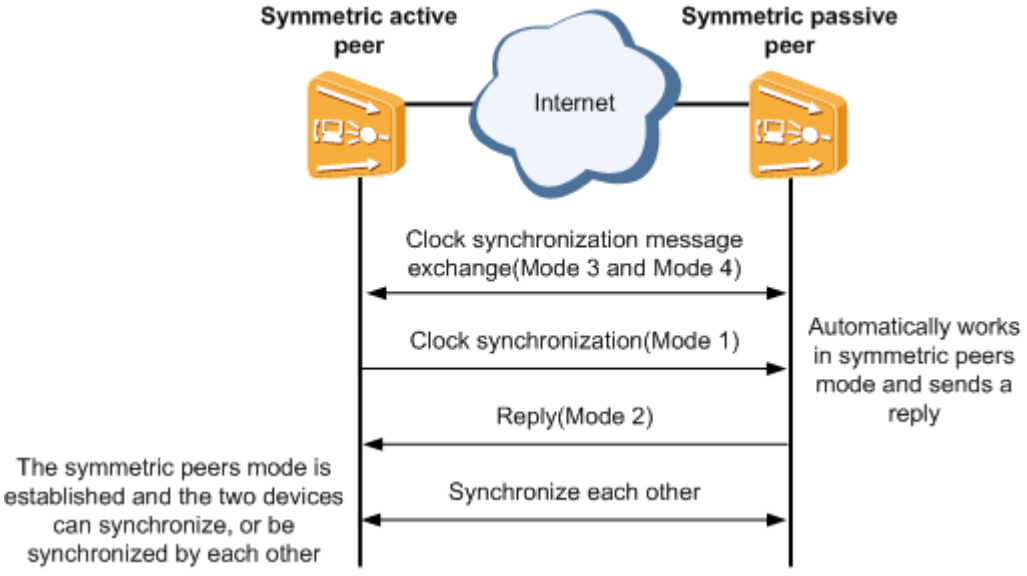
In this mode, the active peer and the passive peer can be synchronized with each other. To be specific, the higher stratum (lower level) peer is synchronized with the lower stratum (higher level) peer. The active and passive peers firstly exchange NTP packets whose values of Mode fields are 3 (sent by the client) and NTP packets whose values are 4 (sent by the server).

- Active peer: A host that functions as an active peer sends packets periodically. The value of the Mode field in a packet is set to 1. This indicates that the packet is sent by an active peer, without considering whether its peer is reachable and which stratum its peer is on. The active peer can provide time information about the local clock for its peer, or synchronize the time information about the local clock based on that of the peer clock.
- Passive peer: A host that functions as a passive peer receives packets from the active peer and sends response packets. The value of the Mode field in a response packet is set to 2. This indicates that the packer is sent by a passive peer. The passive peer can provide time information about the local clock for its peer, or synchronize the time information about the local clock based on that of the peer clock.
- Prerequisites for a host to function as a passive peer: The packets received by the local host are sent by an active peer. The number of the stratum that the active peer is on must be less than or equal to the number of the stratum that the local host is on. In addition, the routes between the local host and the active peer must be reachable.

NOTE

The host operating in passive mode is at the lower stratum in the synchronization subnet. You do not need to obtain information about the peer in advance because the connection between peers is not set up and status variables are not configured unless the passive host receives NTP messages from the peer.

Figure 14-34 Peer mode



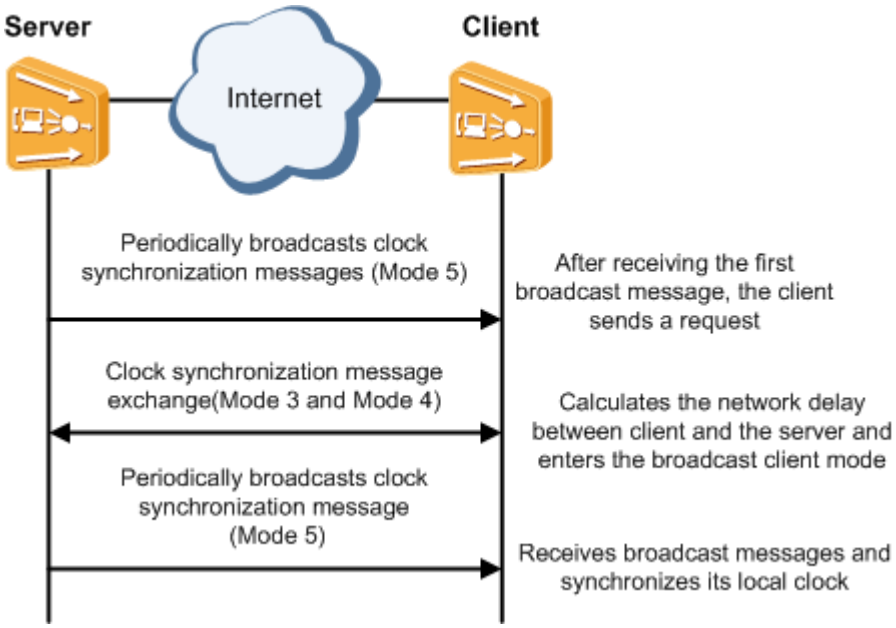
 Copyright © Huawei Technologies Co., Ltd.

Broadcast Mode

- A host that runs in broadcast mode sends clock synchronization packets to the broadcast address 255.255.255.255 periodically. The value of the Mode field in a packet is set to 5. This indicates that the packet is sent by a host that runs in broadcast mode, without considering whether its peer is reachable and which stratum its peer is on. The host running in broadcast mode is usually a time server running high-speed broadcast media on the network, which provides synchronization information for all of its peers but does not alter the clock of its own.
- The client listens to the broadcast packets sent from the server. When the client receives the first broadcast packet, the client and server exchange NTP packets whose values of Mode fields are 3 (sent by the client) and the NTP packets whose values of Mode fields are 4 (sent by the server). In this process, the client enables the server/client mode for a short time to exchange information with the remote server. This allows the client to obtain the network delay between the client and the server. Then, the client returns the broadcast mode, and continues to sense the incoming broadcast packets to synchronize the local clock.

The broadcast mode is applied to the high speed network that has multiple workstations and does not require high accuracy. In a typical scenario, one or more time servers on the network periodically send broadcast packets to the workstations. The delay of packet transmission in a LAN is at the milliseconds level.

Figure 14-35 Broadcast mode



Multicast Mode

- A server running in multicast mode sends clock synchronization packets to a multicast address periodically. The value of the Mode field in a packet is set to 5. This indicates that the packet is sent by a host that runs in multicast mode.

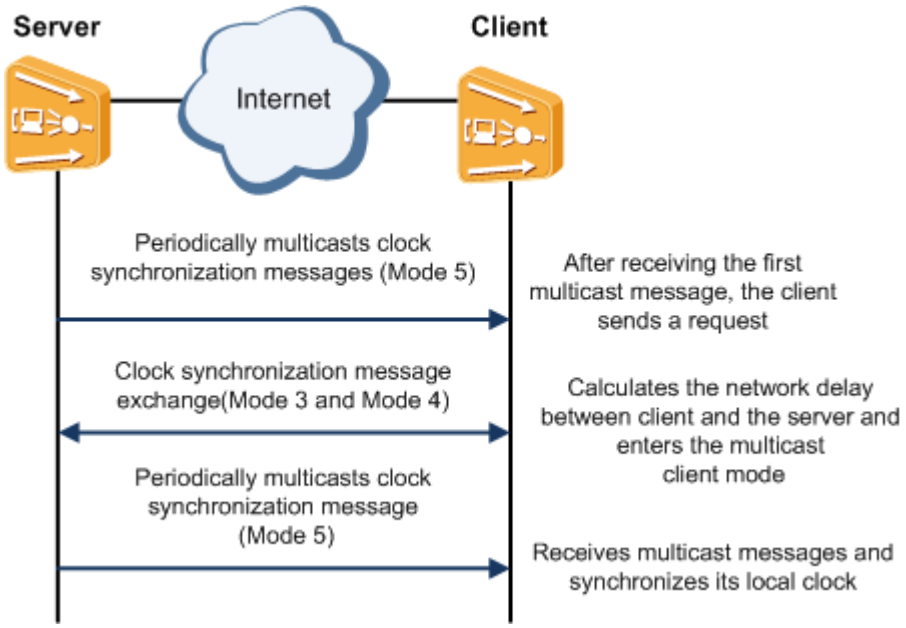
 Copyright © Huawei Technologies Co., Ltd.

The host running in multicast mode is usually a time server running high-speed broadcast media on the network, which provides synchronization information for all of its peers but does not alter the clock of its own.

- The client listens to the multicast packets from the server. When the client receives the first multicast packet, the client and the server exchange NTP packets whose values of Mode fields are 3 (sent by the client) and the NTP packets whose values of Mode fields are 4 (sent by the server). In this process, the client enables the server/client mode for a short time to exchange information with the remote server. This allows the client to obtain the network delay between the client and the server. Then, the client returns the multicast mode, and continues to sense the incoming multicast packets to synchronize the local clock.

Multicast mode is useful when there are large numbers of clients distributed in a network. This normally results in large number of NTP packets in the network. In the multicast mode, a single NTP multicast packet can potentially reach all the clients in the network and thus reduce the control traffic on the network.

Figure 14-36 Multicast mode



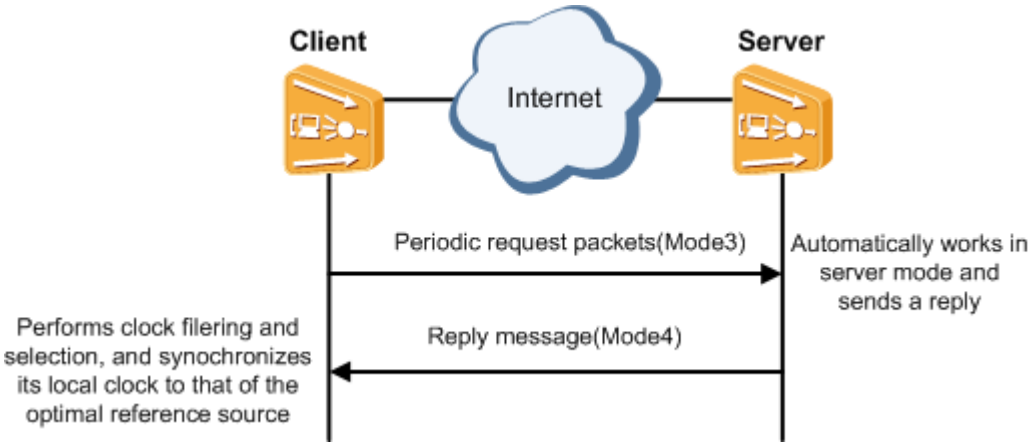
Manycast Mode

In multicast mode, the client first searches for a server by continuously sending NTP packets whose destination IP address is the specified IPv4/IPv6 multicast address. The Mode field in the packets is set to 3 (client mode), and the TTLs of the packets increases from 1. A designated NTP manycast server listens for packets with that address. If a server receives the packets from the client, it returns an ordinary server (mode 4) packet to the client. If client receives the packets from the server, temporary C/S connection between them is established. After a certain period of time, the client stops the search process and select the best NTP server among all connections. Other connections not selected will be aged out.

 Copyright © Huawei Technologies Co., Ltd.

Manycast mode is applied to a small set of servers scattered over the network. Clients can discover and synchronize to the closest manycast server. Manycast can especially be used where the identity of the server is not fixed and a change of server does not require reconfiguration of all the clients in the network.

Figure 14-37 Manycast mode



Kiss-o'-Death (KOD)

KOD packets provide useful information to a client and are used for status reporting and access control. When KOD is enabled at the server, the server may send packets with kiss codes DENY and RATE to the client.

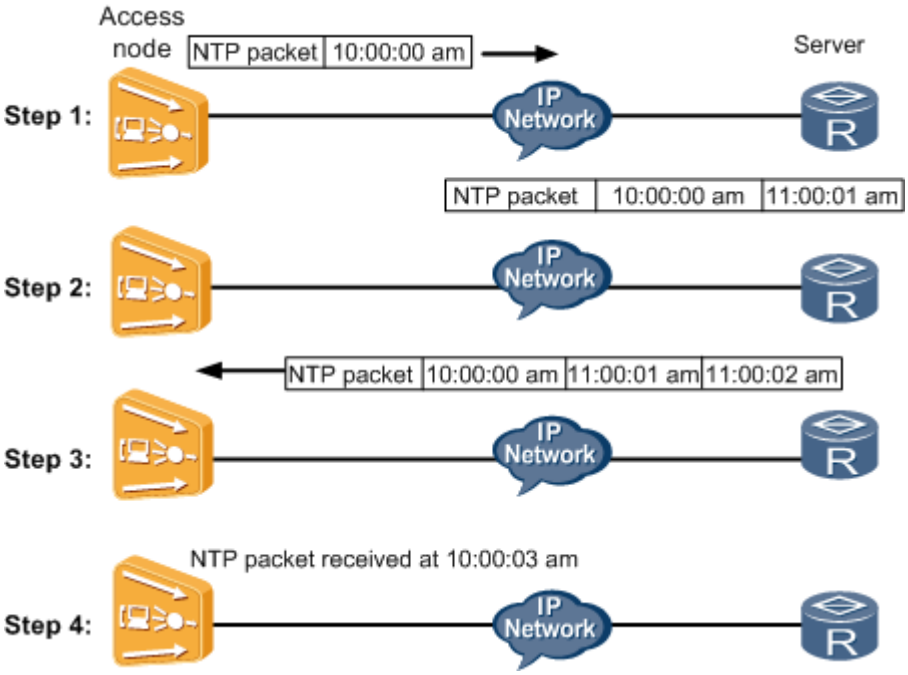
- When the client receives packet with kiss code DENY, the client demobilizes any associations with that server and stops sending packets to that server.
- When the client receives packet with kiss code RATE, the client immediately reduces its polling interval to that server and continues to reduce it each time it receives a RATE kiss code.

14.7.2.3 NTP Working Principle

As shown in **Figure 14-38**, the MA5800 serves as the NTP client and the router serves as the NTP server. The MA5800 uses the time of the router as the reference and synchronizes its time with the router through NTP.

 Copyright © Huawei Technologies Co., Ltd.

Figure 14-38 Operating principle of NTP



1. The MA5800 sends an NTP packet to the router. This packet contains the timestamp when it leaves the MA5800. Assume that the timestamp is 10:00:00 am (T1).
2. When the NTP packet arrives at the router, the router adds its timestamp to the packet. Assume that the timestamp is 11:00:01 am (T2).
3. When the NTP packet leaves the router, the router adds another timestamp to the packet. Assume that the timestamp is 11:00:02 am (T3).
4. When the MA5800 receives the response packet, it adds a new timestamp to the packet. Assume that the timestamp is 10:00:03 am (T4).

Now, the MA5800 has sufficient information to calculate two important parameters:

- The delay for a round trip of the NTP packet=(T4-T1)-(T3-T2).
- Offset between the MA5800 and the router=((T2-T1)-(T4-T3))/2

In this way, the MA5800 can set its clock according to the information and thus keeps its clock synchronized with that of the router.

14.7.2.4 NTP Security Mechanism

When a time server in the subnet is faulty or data is maliciously modified or destroyed, timekeeping on other time servers in the subnet should not be affected. To meet this requirement, NTP provides two security mechanisms: access right and NTP authentication to guarantee the network security.

 Copyright © Huawei Technologies Co., Ltd.

Access Right Control

You can run the **ntp-service access** command to set the right to access the NTP service of the MA5800 so as to protect the NTP service.

The MA5800 supports five levels of access rights, as shown in the following table.

Table 14-12 NTP access rights supported by the MA5800.

Access Right	Description
peer	Indicates the minimum access right. The remote end can perform time requests and control queries for the local NTP service. The local clock can also be synchronized with the clock of the remote server.
server	Indicates that the remote end can perform time requests and control queries for the local NTP service. The local clock, however, cannot be synchronized with the clock of the remote server.
synchronization	Indicates that the remote end can perform time requests only for the local NTP service.
query	Indicates the maximum access right. The remote end can perform control queries only for the local NTP service.
limited	Controls the incoming packet rate and kiss code is sent when KoD is enabled.

NTP Authentication

The NTP authentication can be used on security-critical networks. If the NTP authentication is enabled, the MA5800 uses MD5-algorithm or HMAC-SHA256-algorithm keys to authenticate users who access the NTP service. You are recommended to use the HMAC-SHA256 algorithm to improve security. The NTP authentication must be configured on both the server and the client.

The key on the client must match the one on the server for the user to pass the NTP authentication.

The process for configuring the NTP authentication on the server and the client is as follows:

1. Enable the NTP authentication.
2. Configure the NTP authentication key.
3. Declare that the NTP authentication key is trustworthy.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

- The client is synchronized to only the server that provides the reliable key. If the key provided by the server is unreliable, the client is not synchronized to the server.
- When the client is configured with the NTP authentication, if only the server is configured with same authentication key as the client, the client can pass the authentication. Here, the server does not need to enable the NTP authentication function or declare that the key is reliable.

14.7.3 Configuring the Network Time

Configure the NTP protocol to keep the time of all devices on the network synchronized, so that the MA5800 implements various service applications based on universal time, such as the network management system and the network accounting system.

Context

Introduction to the NTP Protocol:

- The Network Time Protocol (NTP) is an application layer protocol defined in RFC 1305, which is used to synchronize the times of the distributed time server and the client. The RFC defines the structures, arithmetics, entities and protocols used in the implementation of NTP.
- NTP is developed from the time protocol and the ICMP timestamp message protocol, with special design on the aspects of accuracy and robustness.
- NTP runs over UDP with port number as 123.
- Any local system that runs NTP can be time synchronized by other clock sources, and also act as a clock source to synchronize other clocks. In addition, mutual synchronization can be done through NTP packets exchanges.

NTP is applied to the following situations where all the clocks of hosts or routers in a network need to be consistent:

- In the network management, an analysis of log or debugging information collected from different routers needs time for reference.
- The charging system requires the clocks of all devices to be consistent.
- Completing certain functions, for example, timing restart of all the routers in a network requires the clocks of all the routers be consistent.
- When several systems work together on the same complicate event, they have to take the same clock for reference to ensure a correct implementation order.
- Incremental backup between the backup server and clients requires clocks on them be synchronized.

When all the devices on a network need to be synchronized, it is almost impossible for an administrator to manually change the system clock by using a command line. This is because the work load is heavy and clock accuracy cannot be ensured. NTP can quickly synchronize the clocks of network devices and ensure their precision.

There are four NTP modes: server/client, peer, broadcast and multicast modes. The MA5800 supports all these modes.

 Copyright © Huawei Technologies Co., Ltd.

Default Configuration

Table 14-13 provides the default configuration for NTP.

Table 14-13 Default configuration for NTP

Parameter	Default Value
NTP-service authentication function	Disable
NTP-service authentication key	None
Maximum NTP sessions	100
Clock stratum	16

14.7.3.1 (Optional) Configuring NTP Authentication

This topic describes how to configure NTP authentication to improve the network security and prevent unauthorized users from modifying the clock.

Prerequisites

Before configuring the NTP client/server mode, make sure that the network interface and the routing protocol of the device are configured so that the server and the client are reachable to each other at the network layer.

Context

In certain networks that have strict requirements on security, enable NTP authentication when running the NTP protocol. Configuring NTP authentication is classified into configuring NTP authentication on the client and configuring NTP authentication on the server.

Precaution

- If NTP authentication is not enabled on the client, the client can synchronize with the server, regardless of whether NTP authentication is enabled on the server.
- If NTP authentication is enabled, a reliable key should be configured.
- The configuration of the server must be the same as that of the client.
- When NTP authentication is enabled on the client, the client can pass the authentication if the server is configured with the same key as that of the client. In this case, you do not need to enable NTP authentication on the server or declare that the key is reliable.
- The client is synchronized to only the server that provides the reliable key. If the key provided by the server is unreliable, the client is not synchronized to the server.

 Copyright © Huawei Technologies Co., Ltd.

- The flow of configuring NTP authentication is as follows: start->enable NTP authentication->configure the reliable NTP authentication key->declare the reliable key->end.

Procedure

Step 1 Run the **ntp-service authentication enable** command to enable NTP authentication.

Step 2 Run the **ntp-service authentication-keyid** command to set an NTP authentication key.

Step 3 Run the **ntp-service reliable authentication-keyid** command to declare that the key is reliable.

----End

Example

To enable NTP authentication, set the NTP authentication key as **aNiceKey** with the key number 42, and then define key 42 as a reliable key, do as follows:

```
huawei(config)#ntp-service authentication enable
huawei(config)#ntp-service authentication-keyid 42 authentication-mode md5 plain password123
huawei(config)#ntp-service reliable authentication-keyid 42
```

14.7.3.2 Configuring the NTP Broadcast Mode

This topic describes how to configure the MA5800 for clock synchronization in the NTP broadcast mode. After the configuration is completed, the server periodically broadcasts clock synchronization packets through a specified port, and the client listens to the broadcast packets sent from the server and synchronizes the local clock according to the received broadcast packets.

Prerequisites

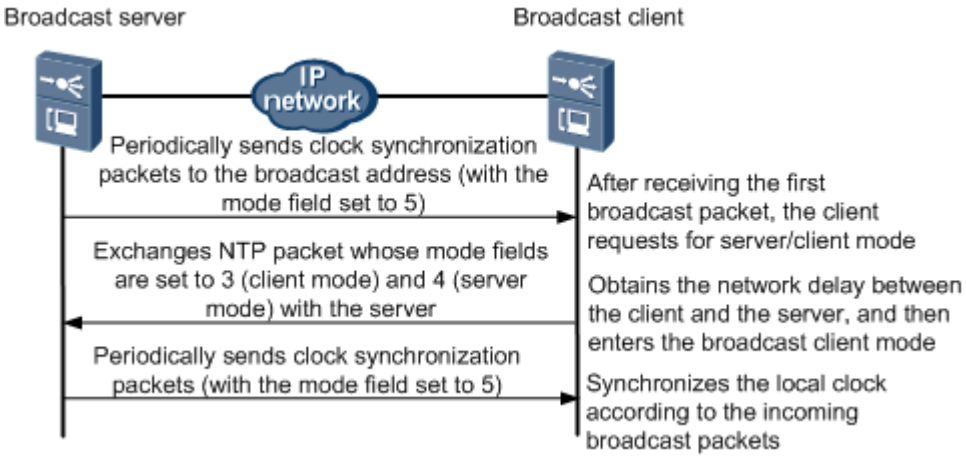
Before configuring the NTP broadcast mode, make sure that the network interface and the routing protocol of the MA5800 are configured so that the server and the client are reachable to each other at the network layer.

Context

In the broadcast mode, the server periodically sends clock synchronization packets to the broadcast address, with the mode field set to 5 (indicating the broadcast mode). The client listens to the broadcast packets sent from the server. After receiving the first broadcast packet, the client exchanges NTP packet whose mode fields are set to 3 (client mode) and 4 (server mode) with the server to estimate the network delay between the client and the server. The client then enters the broadcast client mode, continues to listen to the incoming broadcast packets, and synchronizes the local time according to the incoming broadcast packets, as shown in **Figure 14-39**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 14-39 NTP broadcast mode



Precaution

1. In the broadcast mode, you should configure both the NTP server and the NTP client.
2. The clock stratum of the synchronizing device must be higher than or equal to that of the synchronized device. Otherwise, the clock synchronization fails.

Procedure

- Configure the NTP broadcast server host.
  a. Run the **ntp-service refclock-master** command to configure the local clock as the master NTP clock, and specify the stratum of the master NTP clock.
  b. (Optional) Configure NTP authentication.

     In certain networks that have strict requirements on security, it is recommended that you enable NTP authentication when running the NTP protocol. The configuration of the server must be the same as that of the client.

     i. Run the **ntp-service authentication enable** command to enable NTP authentication.
     ii. Run the **ntp-service authentication-keyid** command to set an NTP authentication key.
     iii. Run the **ntp-service reliable authentication-keyid** command to declare that the key is reliable.
  c. Add a VLAN interface.

     i. Run the **vlan** command to create a VLAN.
     ii. Run the **port vlan** command to add an upstream port to the VLAN so that the user packets carrying the VLAN tag are transmitted upstream through the upstream port.

 Copyright © Huawei Technologies Co., Ltd.

iii. In the global config mode, run the **interface vlan** command to create a VLAN interface, and then enter the VLANIF mode to configure the Layer 3 interface.

iv. Configure an IP address for the VLAN interface based on the actual requirement.

- Run the **ip address** command to configure the IPv4 address and subnet mask of the VLAN interface so that the IP packets in the VLAN can participate in the Layer 3 forwarding.

d. Run the **ntp-service broadcast-server** command to configure the NTP broadcast server mode of the host, and specify the key ID for the server to send packets to the client.

e. Run the **display ntp-service status** command to verify the configuration result.

- Configure the NTP broadcast client host.

a. (Optional) Configure NTP authentication.

In certain networks that have strict requirements on security, it is recommended that you enable NTP authentication when running the NTP protocol. The configuration of the server must be the same as that of the client.

i. Run the **ntp-service authentication enable** command to enable NTP authentication.

ii. Run the **ntp-service authentication-keyid** command to set an NTP authentication key.

iii. Run the **ntp-service reliable authentication-keyid** command to declare that the key is reliable.

b. Add a VLAN interface.

i. Run the **vlan** command to create a VLAN.

ii. Run the **port vlan** command to add an upstream port to the VLAN so that the user packets carrying the VLAN tag are transmitted upstream through the upstream port.

iii. In the global config mode, run the **interface vlan** command to create a VLAN interface, and then enter the VLANIF mode to configure the Layer 3 interface.

iv. Configure an IP address for the VLAN interface based on the actual requirement.

- Run the **ip address** command to configure the IPv4 address and subnet mask of the VLAN interface so that the IP packets in the VLAN can participate in the Layer 3 forwarding.

c. Run the **ntp-service broadcast-client** command to configure a host as the NTP broadcast client.

----End

Example

(IPv4) Assume the following configurations: MA5800_S uses the local clock as the master NTP clock on stratum 2 and works in the NTP broadcast mode,

 Copyright © Huawei Technologies Co., Ltd.

broadcasting clock synchronization packets periodically through IP address 10.10.10.10/24 of the Layer 3 interface of VLAN 2, and MA5800_C functions as the NTP client, listening to the broadcast packets sent from the server through IP address 10.10.10.20/24 of the Layer 3 interface of VLAN 2 and synchronizing with the clock on the broadcast server. To perform these configurations, do as follows:

1. On MA5800_S:

```
huawei(config)#undo ntp-service server disable
huawei(config)#ntp-service refclock-master 2
huawei(config)#vlan 2 standard
huawei(config)#port vlan 2 0/9 1
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.10.10.10 24
huawei(config-if-vlanif2)#ntp-service broadcast-server
huawei(config-if-vlanif2)#quit
```

2. On MA5800_C:

```
huawei(config)#undo ntp-service server disable
huawei(config)#vlan 2 standard
huawei(config)#port vlan 2 0/9 1
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.10.10.20 24
huawei(config-if-vlanif2)#ntp-service broadcast-client
huawei(config-if-vlanif2)#quit
```

14.7.3.3 Configuring the NTP Multicast Mode

This topic describes how to configure the MA5800 for clock synchronization in the NTP multicast mode. After the configuration is completed, the server periodically multicasts clock synchronization packets through a specified port, and the client listens to the multicast packets sent from the server and synchronizes the local clock according to the received multicast packets.

Prerequisites

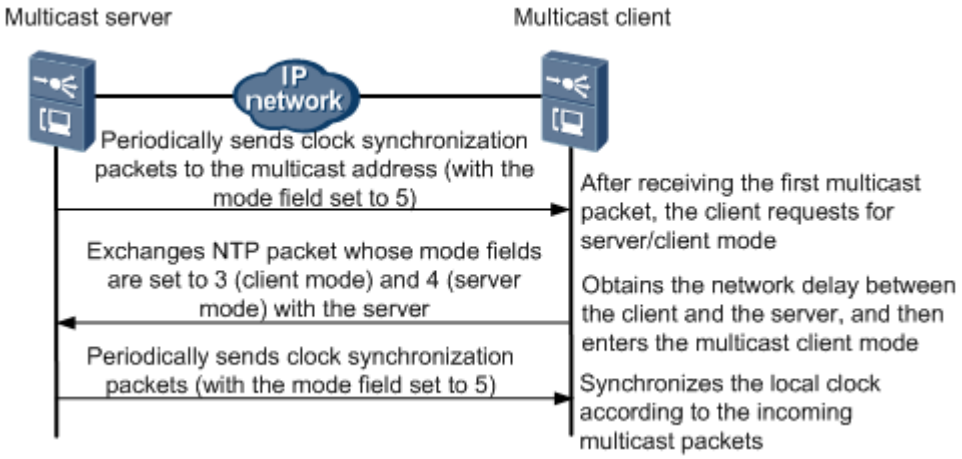
Before configuring the NTP multicast mode, make sure that the network interface and the routing protocol of the MA5800 are configured so that the server and the client are reachable to each other at the network layer.

Context

In the multicast mode, the server periodically sends clock synchronization packets to the multicast address configured by the user. The default NTP multicast address 224.0.1.1 is used if the multicast address is not configured. The mode field of clock synchronization packet is set to 5 (multicast mode). The client listens to the multicast packets sent from the server. After receiving the first multicast packet, the client exchanges NTP packet whose mode fields are set to 3 (client mode) and 4 (server mode) with the server to estimate the network delay between the client and the server. The client then enters the multicast client mode, continues to listen to the incoming multicast packets, and synchronizes the local clock according to the incoming multicast packets, as shown in **Figure 14-40**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 14-40 NTP multicast mode



Precaution

1. In the multicast mode, you should configure both the NTP server and the NTP client.
2. The clock stratum of the synchronizing device must be higher than or equal to that of the synchronized device. Otherwise, the clock synchronization fails.

Procedure

- Configure the NTP multicast server host.
  a. Run the **ntp-service refclock-master** command to configure the local clock as the master NTP clock, and specify the stratum of the master NTP clock.
  b. (Optional) Configure NTP authentication.

     In certain networks that have strict requirements on security, it is recommended that you enable NTP authentication when running the NTP protocol. The configuration of the server must be the same as that of the client.

     i. Run the **ntp-service authentication enable** command to enable NTP authentication.
     ii. Run the **ntp-service authentication-keyid** command to set an NTP authentication key.
     iii. Run the **ntp-service reliable authentication-keyid** command to declare that the key is reliable.
  c. Add a VLAN interface.

     i. Run the **vlan** command to create a VLAN.
     ii. Run the **port vlan** command to add an upstream port to the VLAN so that the user packets carrying the VLAN tag are transmitted upstream through the upstream port.

 Copyright © Huawei Technologies Co., Ltd.

iii. In the global config mode, run the **interface vlan** command to create a VLAN interface, and then enter the VLANIF mode to configure the Layer 3 interface.

iv. Configure the IP address of the VLAN interface.

- Run the **ip address** command to configure the IPv4 address and subnet mask of the VLAN interface so that the IP packets in the VLAN can participate in the Layer 3 forwarding.

d. Run the **ntp-service multicast-server** command to configure the NTP multicast server mode of the host, and specify the key ID for the server to send packets to the client.

- Configure the NTP multicast client host.

a. (Optional) Configure NTP authentication.

In certain networks that have strict requirements on security, it is recommended that you enable NTP authentication when running the NTP protocol. The configuration of the server must be the same as that of the client.

i. Run the **ntp-service authentication enable** command to enable NTP authentication.

ii. Run the **ntp-service authentication-keyid** command to set an NTP authentication key.

iii. Run the **ntp-service reliable authentication-keyid** command to declare that the key is reliable.

b. Add a VLAN interface.

i. Run the **vlan** command to create a VLAN.

ii. Run the **port vlan** command to add an upstream port to the VLAN so that the user packets carrying the VLAN tag are transmitted upstream through the upstream port.

iii. In the global config mode, run the **interface vlan** command to create a VLAN interface, and then enter the VLANIF mode to configure the Layer 3 interface.

iv. Configure the IP address of the VLAN interface.

- Run the **ip address** command to configure the IPv4 address and subnet mask of the VLAN interface so that the IP packets in the VLAN can participate in the Layer 3 forwarding.

c. Run the **ntp-service multicast-client** command to configure a host as the NTP multicast client.

----End

Example

(IPv4) Assume the following configurations: MA5800_S uses the local clock as the master NTP clock on stratum 2 and works in the NTP multicast mode, multicasting clock synchronization packets periodically through IP address 10.10.10.10/24 of the Layer 3 interface of VLAN 2, and MA5800_C functions as the NTP client, listening to the multicast packets sent from the server through IP address 10.10.10.20/24 of the Layer 3 interface of VLAN 2 and synchronizing with the clock on the multicast server. To perform these configurations, do as follows:

 Copyright © Huawei Technologies Co., Ltd.

1. On MA5800_S:

```
huawei(config)#ntp-service refclock-master 2
huawei(config)#vlan 2 standard
huawei(config)#port vlan 2 0/9 1
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.10.10.10 24
huawei(config-if-vlanif2)#ntp-service multicast-server
huawei(config-if-vlanif2)#quit
```

2. On MA5800_C:

```
huawei(config)#vlan 2 standard
huawei(config)#port vlan 2 0/9 1
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.10.10.20 24
huawei(config-if-vlanif2)#ntp-service multicast-client
huawei(config-if-vlanif2)#quit
```

14.7.3.4 Configuring the Unicast NTP Client

This topic describes how to configure the MA5800 as the NTP client to synchronize with the NTP server in the network.

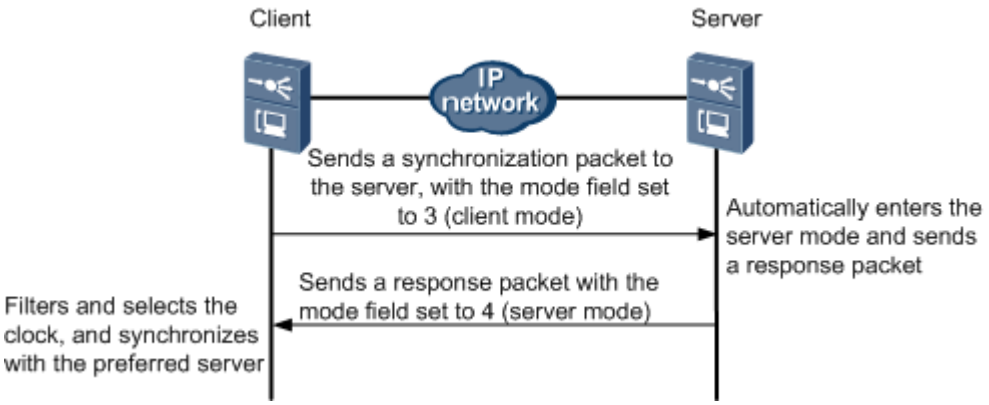
Prerequisites

Before configuring the NTP client/server mode, make sure that the network interface and the routing protocol of the MA5800 are configured so that the server and the client are reachable to each other at the network layer.

Context

In the client/server mode, the client sends a synchronization packet to the server, with the mode field set to 3 (client mode). After receiving the packet, the server automatically enters the server mode and sends a response packet with the mode field set to 4 (server mode). After receiving the response from the server, the client filters and selects the time, and synchronizes with the preferred server, as shown in **Figure 14-41**.

Figure 14-41 NTP client/server mode



Precaution

1. In the client/server mode, you need to configure only the client, and do not need to configure the server.

 Copyright © Huawei Technologies Co., Ltd.

2. The clock stratum of the synchronizing device must be lower than or equal to that of the synchronized device. Otherwise, the clock synchronization fails.

Procedure

Step 1 Add a VLAN interface.

1. Run the **vlan** command to create a VLAN.
2. Run the **port vlan** command to add an upstream port to the VLAN so that the user packets carrying the VLAN tag are transmitted upstream through the upstream port.
3. In the global config mode, run the **interface vlan** command to create a VLAN interface, and then enter the VLANIF mode to configure the Layer 3 interface.
4. Configure the IP address of the VLAN interface.
   - Run the **ip address** command to configure the IPv4 address and subnet mask of the VLAN interface so that the IP packets in the VLAN can participate in the Layer 3 forwarding.

Step 2 Run the **ntp-service unicast-server** command on the client to configure the NTP unicast server mode, and specify the IP address of the remote server that functions as the local timer server and the interface for transmitting and receiving NTP packets.

NOTE

- In this command, *ip-address* is a unicast address, which cannot be a broadcast address, a multicast address, or the IP address of a local clock.
- After the source interface of the NTP packets is specified by *source-interface*, the source IP address of the NTP packets is configured as the primary IP address of the specified interface.
- A server can function as a time server to synchronize other devices only after its clock is synchronized.
- When the clock stratum of the server is higher than or equal to that of the client, the client does not synchronize with the server.

You can run the **ntp-service unicast-server** command for multiple times to configure multiple servers. Then, the client selects the best server according to clock priorities or preemption.

For example, assuming that there are two servers A and B. A sets up a temporary NTP session, and then B sets up another. When the conditions (level and precision) of A and B are the same, **Table 14-14** shows the clock source selection results.

Table 14-14 Clock source selection results

Mode (A)	Mode (B)	Clock Source
None	None	A
None	Priority	B
None	Preemption	A
Priority	None	A

 Copyright © Huawei Technologies Co., Ltd.

Mode (A)	Mode (B)	Clock Source
Priority	Priority	A
Priority	Preemption	A
Preemption	None	B
Preemption	Priority	B
Preemption	Preemption	A

Step 3 (Optional) Configure the ACL rules.

Filter the packets that pass through the Layer 3 interface. Only the IP packet from the clock server is allowed to access the Layer 3 interface. Other unauthorized packets are not allowed to access the Layer 3 interface. It is recommended to use the ACL rules for the system that has high requirements on security.

1. Run the **acl adv-acl-numbe** command to create an ACL for the IPv4 system.
2. Run the **rule** command to classify traffic according to the source IP address, destination IP address, type of the protocol over IP, and features or protocol of the packet, allowing or forbidding the data packets that meet related conditions to pass.
3. Run the **packet-filter** command to configure an ACL filtering rule for a specified port, and make the configuration take effect.

----End

Example

(IPv4)Assume the following configurations : One MA5800 functions as the NTP server (IP address: 10.20.20.20/24), the other MA5800 (IP address of the Layer 3 interface of VLAN 2: 10.10.10.10/24, gateway IP address: 10.10.10.1) functions as the NTP client, the NTP client sends the clock synchronization request packet through the VLAN interface to the NTP server, the NTP server responds to the request packet, and ACL rules are configured to allow only IP packets from the clock server to access the Layer 3 interface. To perform these configurations, do as follows:

```
huawei(config)#vlan 2 standard
huawei(config)#port vlan 2 0/9 1
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.10.10.10 24
huawei(config-if-vlanif2)#quit
huawei(config)#undo ntp-service server disable
huawei(config)#ntp-service unicast-server 10.20.20.20 source-interface vlanif 2
huawei(config)#acl 3010
huawei(config-acl-adv-3010)#rule deny ip source any destination 10.10.10.10
 0.0.0.0
huawei(config-acl-adv-3010)#rule permit ip source 10.20.20.20 0.0.0.0 destination
 10.10.10.10 0.0.0.0
huawei(config-acl-adv-3010)#quit
huawei(config)#packet-filter inbound ip-group 3010 port 0/9/1
```

14.7.3.5 Configuring the NTP Peer

This topic describes how to configure the MA5800 for clock synchronization in the NTP peer mode. In the peer mode, configure only the active peer, and the passive

 Copyright © Huawei Technologies Co., Ltd.

peer does not need to be configured. In the peer mode, the active peer and the passive peer can synchronize with each other. The peer with a higher clock stratum is synchronized by the peer with a lower clock stratum.

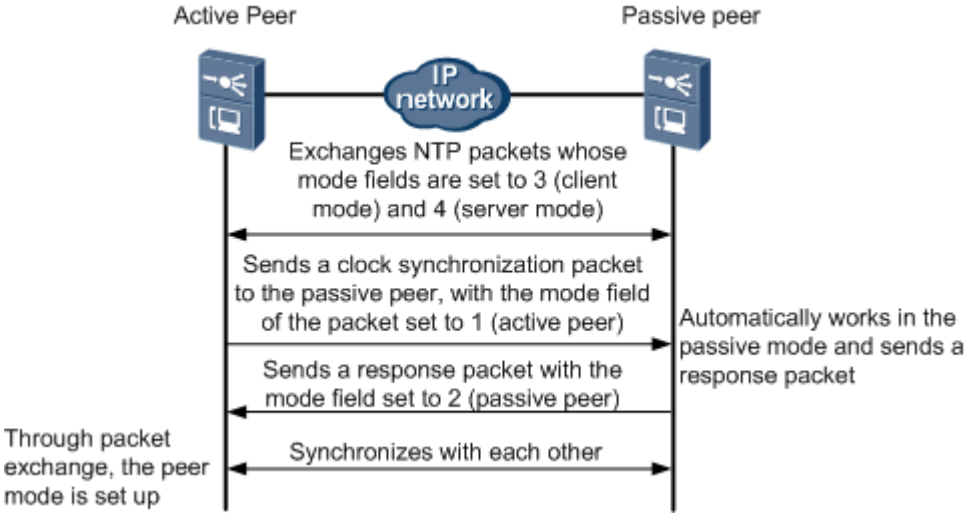
Prerequisites

Before configuring the NTP peer mode, make sure that the network interface and the routing protocol of the MA5800 are configured so that the server and the client are reachable to each other at the network layer.

Context

In the peer mode, the active peer and the passive peer exchange NTP packets whose mode fields are set to 3 (client mode) and 4 (server mode). Then, the active peer sends a clock synchronization packet to the passive peer, with the mode field of the packet set to 1 (active peer). After receiving the packet, the passive peer automatically works in the passive mode and sends a response packet with the mode field set to 2 (passive peer). Through packet exchange, the peer mode is set up. The active peer and the passive peer can synchronize with each other. If both the clock of the active peer and that of the passive peer are synchronized, the clock on a lower stratum is used, as shown in **Figure 14-42**.

Figure 14-42 NTP peer mode



Precaution

1. In the peer mode, you need to configure the NTP mode only on the active peer.
2. The peers determine clock synchronization according to the clock stratum instead of according to whether the peer is an active peer.

Procedure

Step 1 Enable the NTP server functionality.

 Copyright © Huawei Technologies Co., Ltd.

1. Run the **undo ntp-service server disable** command to enable IPv4 NTP server functionality.

Step 2 Configure the NTP active peer.

1. Run the **ntp-service refclock-master** command to configure the local clock as the master NTP clock, and specify the stratum of the master NTP clock.
2. Run the **ntp-service unicast-peer** command to configure the NTP peer mode, and specify the IP address of the remote server that functions as the local timer server and the interface for transmitting and receiving NTP packets.

   NOTE

   - In this command, *ip-address* is a unicast address, which cannot be a broadcast address, a multicast address, or the IP address of a reference clock.
   - After the source interface of the NTP packets is specified by *source-interface*, the source IP address of the NTP packets is configured as the primary IP address of the specified interface.

Step 3 Add a VLAN interface.

1. Run the **vlan** command to create a VLAN.
2. Run the **port vlan** command to add an upstream port to the VLAN so that the user packets carrying the VLAN tag are transmitted upstream through the upstream port.
3. In the global config mode, run the **interface vlan** command to create a VLAN interface, and then enter the VLANIF mode to configure the Layer 3 interface.
4. Configure the IP address based on the system.
   - Run the **ip address** command to configure the IPv4 address and subnet mask of the VLAN interface so that the IP packets in the VLAN can participate in the Layer 3 forwarding.

----End

Example

Assume the following configurations (IPv4): One MA5800 functions as the NTP active peer (IP address of the Layer 3 interface of VLAN 2: 10.10.10.10/24) and works on clock stratum 4, the other MA5800 (IP address: 10.10.10.20/24) functions as the NTP passive peer, the active peer sends a clock synchronization request packet through the VLAN interface to the passive peer, the passive peer responds to the request packet, and the peer with a higher clock stratum is synchronized by the peer with a lower clock stratum. To perform these configurations, do as follows:

```
huawei(config)#undo ntp-service server disable
huawei(config)#ntp-service refclock-master 4
huawei(config)#ntp-service unicast-peer 10.10.10.20 source-interface vlanif 2
huawei(config)#vlan 2 standard
huawei(config)#port vlan 2 0/9 1
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.10.10.10 24
huawei(config-if-vlanif2)#quit
```

14.7.3.6 Configuring the NTP Multicast Mode

This topic describes how to configure the MA5800 for clock synchronization in the NTP multicast mode. The multicast mode is a mechanism used by NTP clients to

 Copyright © Huawei Technologies Co., Ltd.

dynamically discover NTP servers. After the multicast mode is enabled, unicast NTP servers do not need to be manually configured for clients.

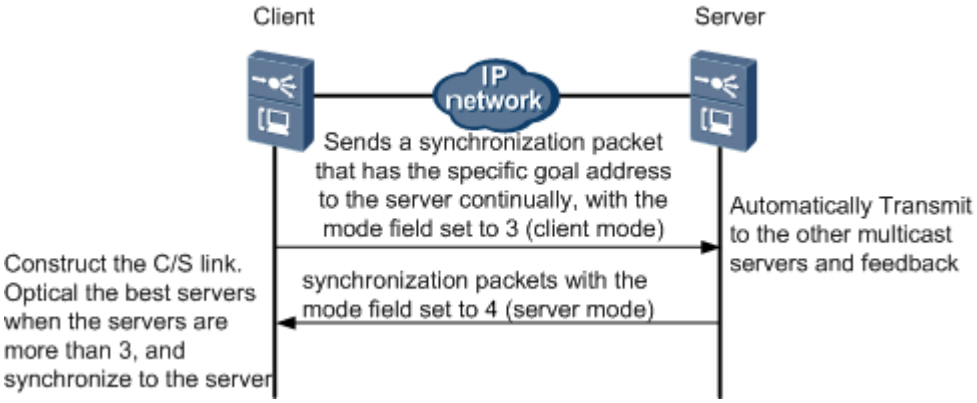
Prerequisites

Before configuring the NTP multicast mode, make sure that the network interface and the routing protocol of the MA5800 are configured so that the server and the client are reachable to each other at the network layer.

Context

In the multicast mode, multicast clients search multicast NTP servers by sending NTP packets using the specified multicast IP address as the destination IP address. The mode field in the packets is set to 3 (client mode). Multicast NTP servers listen to the multicast IP address, and send response packets after receiving the packets. The mode field in the response packets is set to 4 (server mode). After receiving the response packets, the multicast clients set up temporary client/server (C/S) connections. After setting up a certain number of C/S connections (the required number of C/S connections have a fixed value of 3), the multicast clients stop searching and select the optimal NTP server from the C/S connections. The unselected C/S connections will be aged. **Figure 14-43** shows the NTP multicast mode.

Figure 14-43 NTP multicast mode



Precaution

1. In the multicast mode, you should configure both the NTP server and the NTP client.
2. The clock stratum of the synchronizing device must be higher than or equal to that of the synchronized device. Otherwise, the clock synchronization fails.

Procedure

- Configure the NTP multicast server host.
  a. Run the **undo ntp-service server disable** command to enable the IPv4 NTP server function.

 Copyright © Huawei Technologies Co., Ltd.

b. Run the **ntp-service refclock-master** command to configure the local clock as the master NTP clock, and specify the stratum of the master NTP clock.

c. (Optional) Configure NTP authentication.

In certain networks that have strict requirements on security, it is recommended that you enable NTP authentication when running the NTP protocol. The configuration of the server must be the same as that of the client.

i. Run the **ntp-service authentication enable** command to enable NTP authentication.

ii. Run the **ntp-service authentication-keyid** command to set an NTP authentication key.

iii. Run the **ntp-service reliable authentication-keyid** command to declare that the key is reliable.

d. Add a VLAN interface.

i. Run the **vlan** command to create a VLAN.

ii. Run the **port vlan** command to add an upstream port to the VLAN so that the user packets carrying the VLAN tag are transmitted upstream through the upstream port.

iii. In the global config mode, run the **interface vlan** command to create a VLAN interface, and then enter the VLANIF mode to configure the Layer 3 interface.

iv. Run the **ip address** command to configure the IPv4 address and subnet mask of the VLAN interface so that the IP packets in the VLAN can participate in the Layer 3 forwarding.

e. Run the **ntp-service manycast-server** command to configure the NTP multicast server mode of the host, and specify the key ID for the server to send packets to the client.

- Configure the NTP multicast client host.

a. (Optional) Configure NTP authentication.

In certain networks that have strict requirements on security, it is recommended that you enable NTP authentication when running the NTP protocol. The configuration of the server must be the same as that of the client.

i. Run the **ntp-service authentication enable** command to enable NTP authentication.

ii. Run the **ntp-service authentication-keyid** command to set an NTP authentication key.

iii. Run the **ntp-service reliable authentication-keyid** command to declare that the key is reliable.

b. Add a VLAN interface.

i. Run the **vlan** command to create a VLAN.

ii. Run the **port vlan** command to add an upstream port to the VLAN so that the user packets carrying the VLAN tag are transmitted upstream through the upstream port.

 Copyright © Huawei Technologies Co., Ltd.

iii. In the global config mode, run the **interface vlan** command to create a VLAN interface, and then enter the VLANIF mode to configure the Layer 3 interface.

iv. Run the **ip address** command to configure the IPv4 address and subnet mask of the VLAN interface so that the IP packets in the VLAN can participate in the Layer 3 forwarding.

c. Run the **ntp-service manycast-client** command to configure a host as the NTP multicast client.

----End

Example

(IPv4) Assume the following configurations: MA5800_S uses the local clock as the master NTP clock on stratum 2 and works in the NTP multicast mode, listening to clock synchronization packets that use 10.10.10.10/24 as the destination IP address through IP address 10.10.10.10/24 of the Layer 3 interface of VLAN 2 and establishing C/S connections with multicast clients, and MA5800_C functions as the NTP multicast client, multicasting clock synchronization packets that use 10.10.10.10/24 as the destination IP address continuously through IP address 10.10.10.20/24 of the Layer 3 interface of VLAN 2 and synchronizing with the clock on the multicast server by establishing C/S connections with the multicast server. To perform these configurations, do as follows:

1. On MA5800_S:

```
huawei(config)#undo ntp-service server disable
huawei(config)#ntp-service refclock-master 2
huawei(config)#vlan 2 standard
huawei(config)#port vlan 2 0/9 1
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.10.10.10 24
huawei(config-if-vlanif2)#ntp-service manycast-server
huawei(config-if-vlanif2)#quit
```

2. On MA5800_C:

```
huawei(config)#undo ntp-service server disable
huawei(config)#vlan 2 standard
huawei(config)#port vlan 2 0/9 1
huawei(config)#interface vlanif 2
huawei(config-if-vlanif2)#ip address 10.10.10.20 24
huawei(config-if-vlanif2)#ntp-service manycast-client
huawei(config-if-vlanif2)#quit
```

14.7.3.7 (Optional) Configuring NTPv4 Control Signaling KOD

Kiss-o'-Death (KOD) packets are NTPv4-defined control signaling packets used for state advertisement during Network Time Protocol (NTP) message exchange and for access control. The KOD packets do not apply to time synchronization. The KOD packets can help the users to control the interval that the client transmits packets to the server.

Prerequisites

- The client sends NTP packets to the server for multiple times.
- The server filters packets sent by the client.

 Copyright © Huawei Technologies Co., Ltd.

Context

KOD packets carry KOD codes that are indicated by a four-character ASCII character string. The KOD codes are called KISS codes. Different KISS codes have different meanings. The system supports RATE and DENY KISS codes.

KISS Code	Server Configuration	Application Scenario
DENY	Run the **ntp-service access** command to configure access control. Then the server sends KOD packets that carry DENY KISS codes to the client when the client's access request is denied because of restricted rights.	This configuration applies when an NTP server is only configured for specified clients, which provides protection for the clients.
RATE	Run the **ntp-service access limited** command to configure rate control, and the **ntp-service discard** command to configure the minimum interval and average interval allowed by the server for the client to transmit packets. • When the client sends packets at an interval shorter than the minimum or average interval, the server sends KOD packets that carry RATE KISS codes to the client. • When the client sends packets at an interval longer than the minimum and average intervals, the server does not send any KOD packets.	This configuration applies to servers of poor performance.

Precaution

The following operations are performed only on servers.

 Copyright © Huawei Technologies Co., Ltd.

Procedure

Step 1 Run the **undo ntp-service server disable** command to enable IPv4 NTP server functionality.

Step 2 Run the **ntp-service kod-enable** command to enable the KOD function globally. Then the server sends KOD packets that carry appropriate KISS codes in different scenarios, which controls packet transmission of the client.

Step 3 Run the **ntp-service access** command to enable the access control function on the server. Then the server can control the client's access based on the client's rights. If the client's access is denied, the server sends KOD packets that carry **DENY** KISS codes.

Step 4 Run the **ntp-service discard** command to configure the minimum interval and average interval allowed by the server for the client to transmit packets. Then the server controls packet transmission of the client using **RATE** KISS codes in KOD packets based on packet transmission frequency.

----End

Example

IPv4: Assume that server A in an IPv4 system has poor performance. The access control function needs to be enabled for server A to only receive NTP packets sent by clients at the minimum and average intervals longer than 3. To perform these configurations, do as follows:

```
huawei(config)#undo ntp-service server disable
huawei(config)#ntp-service kod-enable
huawei(config)#ntp-service access limited 2000
huawei(config)#ntp-service discard ntp-service discard min-interval 3 avg-interval 3
```

14.7.4 NTP Standards and Protocols Compliance

Reference documents:

- RFC 1305: Basis of the NTP module requirements specification
- RFC 5905: NTP version 4 Protocol and algorithm specification

 Copyright © Huawei Technologies Co., Ltd.

15 QoS

Quality of service (QoS) is a mechanism to ensure that user requirements for bandwidth, latency, jitter, and packet loss rate are met.

15.1 Introduction to QoS

15.2 QoS Models

15.3 QoS Scheme

15.4 QoS Processing

15.5 Traffic Classification
Traffic classification differentiates services by packet classification according to the characteristics of user Ethernet packets and certain rules, to implement different processing operations and provide different services.

15.6 Priority Marking
According to different priority marking policies, the inner and outer VLAN priorities can be set for service-ports, or the user-side priority can be copied for service-ports.

15.7 Traffic Policing
Before service providers provide subscribers with specific services, a service level agreement (SLA) is generally assigned, in which all service parameters are defined. To ensure that the user traffic can meet the SLA, the user traffic must be policed.

15.8 Traffic Shaping

15.9 Congestion avoidance

15.10 Congestion Management

15.11 ACL
Using the preset access control list (ACL) policy, the system permits or refuses data packets to pass.

15.12 ACLv6

15.13 HQoS
Traditional quality of service (QoS) schedules traffic based on port. Hierarchical quality of service (HQoS) not only controls port traffic at a finer service granularity but also schedules traffic based on service priorities.

 Copyright © Huawei Technologies Co., Ltd.

15.14 End-to-End QoS

15.1 Introduction to QoS

Definition

Quality of service (QoS) is a mechanism that guarantees an expected service level with respect to bandwidth, latency, jitter, and packet loss in a communication network. The following indicators are used to measure QoS:

- Bandwidth: theoretical transmission capacities of a connection.
- Latency: the time required for information to travel from one network node to another network node. A high latency impacts the quality of real-time services (for example, IP telephone service).
- Jitter: variations in latency. Jitter can severely affect the quality of multimedia services (for example, VoD).
- Packet loss rate: the percentage of lost packets to the total packets during network transmission.

Purpose

QoS achieves the following purposes:

- Provides users with assured bandwidth.
- Regulates and controls IP network traffic.
- Reduces packet loss rate.
- Specifies packet priorities.
- Avoids and manages network congestion.
- Provides differentiated services for users.

15.2 QoS Models

QoS Models

The following table describes three types of QoS models that may be used in a network. The QoS model implemented on the MA5800 is the differentiated service (DiffServ) model. The following topics use the DiffServ model to describe QoS unless otherwise specified.

 Copyright © Huawei Technologies Co., Ltd.

Type	Feature	Application
Best-effort service model	In this model, the network forwards data at best-effort rates and no guarantees are provided. The network drops data after all bandwidth is exhausted. This is a simple and unitary service model. It is the default service model for IP networks.	This service model applies to most data services, for example, Email services.
Integrated service model	This model is based on resource reservation. In this model, applications instruct a network to reserve bandwidth using the Resource Reservation Protocol (RSVP), and each unit in the network has to reserve bandwidth for specific data streams.	This service model is not widely used because of the following limitations: ● End-to-end support is required for RSVP. ● RSVP is not highly extendible. ● RSVP protocol packets require large overhead.
DiffServ model	This model is based on priorities. In this model, the network identifies each data stream and provides corresponding QoS guarantees for the data streams. Traffic classification and priority marking are the prerequisites for using this model.	This model involves simple packet processing and is highly extendible. It is used for the following services: ● VoD ● Streaming media ● VoIP ● Video conferencing ● Private-line services

QoS Components in the DiffServ Model

The DiffServ model uses four QoS components, which are described in the following table. You can design holistic QoS policies through flexible combinations of these components.

 Copyright © Huawei Technologies Co., Ltd.

Component	Description
Traffic classification and priority marking	DiffServ operates on the principle of traffic classification and priority marking. Traffic classification: Data packets are placed into different traffic classes. Traffic classification does not modify the data packets. Priority marking: Data packets of each traffic class are marked with a specific priority. Priority marking modifies the data packets because the marking changes the values of some packet fields.
Traffic policing and shaping	Before providing services for a subscriber, service providers usually sign a service level agreement (SLA) with the subscriber to define the level of service. Traffic policing: Packets exceeding the SLA will be dropped. Traffic shaping: Packets exceeding the SLA are buffered and the transmission of these packets is resumed when bandwidth is sufficient.
Congestion management	Congestion management controls the sequence of data transmission when congestion occurs in a network.
Congestion avoidance	Congestion avoidance is a traffic control mechanism that actively drops packets when congestion occurs in a network. This mechanism addresses network overload by regulating the network traffic.

15.3 QoS Scheme

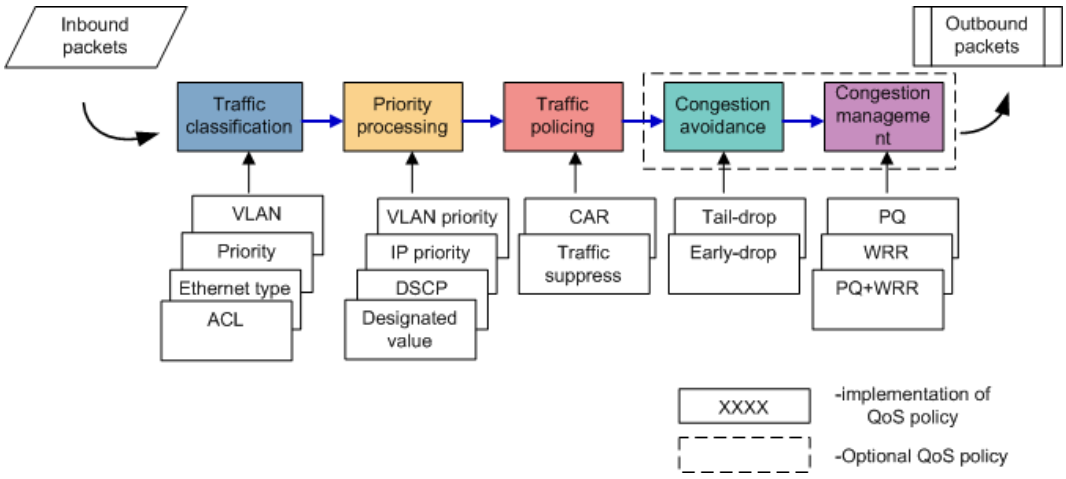
QoS Scheme

The MA5800 use the differentiated service (DiffServ) model. **Figure 15-1** shows the QoS scheme of this model. The QoS scheme works as follows:

- Before forwarding the packets through an outbound interface, QoS policies, such as traffic classification, priority processing, traffic policing, congestion avoidance, and congestion management, are implemented on the packets.
- Congestion avoidance and congestion management are optional. They are not required if no congestion occurs (when the backplane bandwidth is higher than or equal to the maximum bandwidth actually required by a board).
- Each QoS policy can be implemented using multiple techniques. For details, see description about the QoS policies.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-1 QoS scheme

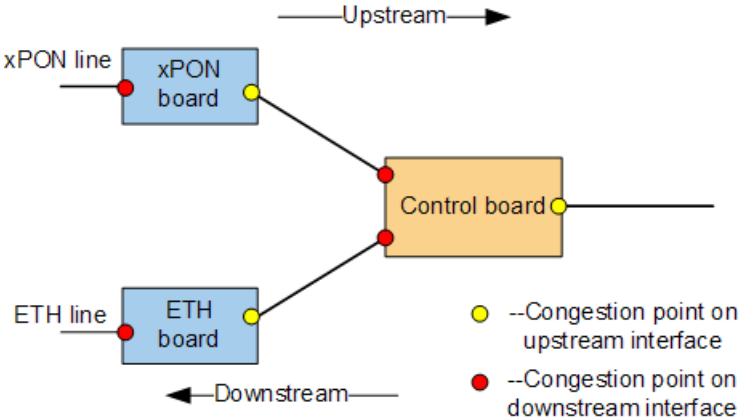


QoS Congestion Point

A network point where traffic congestion is likely to occur is defined as a QoS congestion point. Congestion avoidance and congestion management must be applied to such a point. As shown in **Figure 15-2**, there are QoS congestion points on both the control boards and service boards.

- There is a QoS congestion point on each service board in the downstream direction.
- In the upstream direction of a service board, congestion control is applied depending on the actual demand on bandwidth.

Figure 15-2 Upstream and downstream QoS congestion points



QoS Policies for Service Boards and Control Boards

The following table provides the QoS policies for service boards and control boards.

 Copyright © Huawei Technologies Co., Ltd.

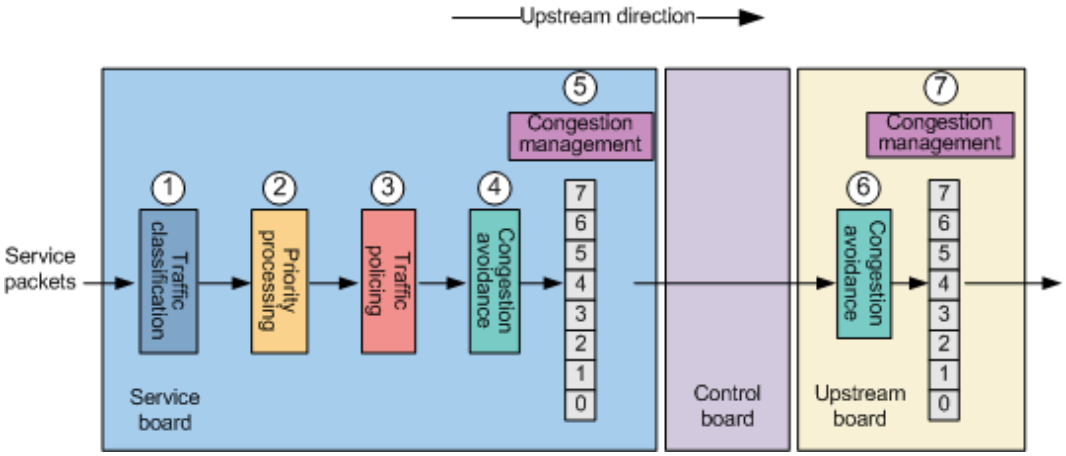
Board Type	QoS Policy	Remarks
Service board	Traffic classification Priority processing Traffic policing Congestion avoidance Congestion management	• Traffic classification on a service board is performed only in the upstream direction. • Congestion avoidance and congestion management are performed on an outbound interface of a service board.
Control board	Traffic classification (based on ACL) Priority processing (based on ACL) Traffic policing (based on ACL) Congestion avoidance Congestion management	Congestion avoidance and congestion management are performed on an outbound interface of a control board. For traffic classification, priority processing, and traffic policing based on ACL on the control board, see "**15.11 ACL**".

15.4 QoS Processing

General QoS Processing - Upstream Direction

Figure 15-3 shows the general QoS processing in the upstream direction.

Figure 15-3 QoS processing in the upstream direction



After the user packets enter a service board from a user port, the service board implements QoS processing on the packets. The following table describes the QoS processing.

 Copyright © Huawei Technologies Co., Ltd.

Step	QoS Policy	Example
1	Traffic classification: The service board places the Ethernet user packets into different classes based on packet characteristics and predefined QoS rules, and services the traffic differently.	Plan different VLANs or priorities for Internet, voice, and IPTV services to distinguish them, because these services have different QoS requirements.
2	Priority processing: When congestion occurs on the local device or the upper-layer network, the service board marks or remarks priorities for the user packets so that the packets are scheduled based on their priorities.	Assign priorities 0, 5, and 4 for Internet, voice, and IPTV services respectively.
3	Traffic policing: The service board implements traffic policing to limit the traffic and burst size of the user packets destined for a connection on the network. When the packets meet certain conditions, for example, when the traffic destined for the connection is too heavy, the service board takes actions accordingly, such as dropping or coloring (resetting the priorities) the packets. Traffic policing enables a port to work at a stable rate, preventing impact on lower-layer devices.	For Internet services, traffic policing provides a committed information rate (CIR) of 8 Mbit/s and limits the peak information rate (PIR) to 10 Mbit/s. When the rate of the Internet service packets is between the CIR and PIR, the service board marks the packets yellow. When the rate exceeds the PIR, the service board drops the packets.
4	Congestion avoidance: When congestion occurs during enqueuing on an outbound interface, the service board drops non-conformant packets to avoid further congestion.	To promptly drop packets based on priority, set the early drop threshold to 30% for 0-priority Internet services. After this setting takes effect, if 30% of the packets that arrive at a queue are Internet service packets, subsequent Internet service packets will be dropped.
5	Congestion management: When congestion occurs during dequeuing on an outbound interface, the service board uses queuing mechanisms to provide QoS guarantees to high-priority packets preferentially.	Use the strict priority (SP) scheduling algorithm to ensure that voice services with priority 5 are scheduled preferentially and Internet services with priority 0 are scheduled last when congestion occurs.

Packets on the control board are not processed for QoS but are processed for only switching.

 Copyright © Huawei Technologies Co., Ltd.

After the packets enter the upstream board, the upstream board implements the following QoS processing.

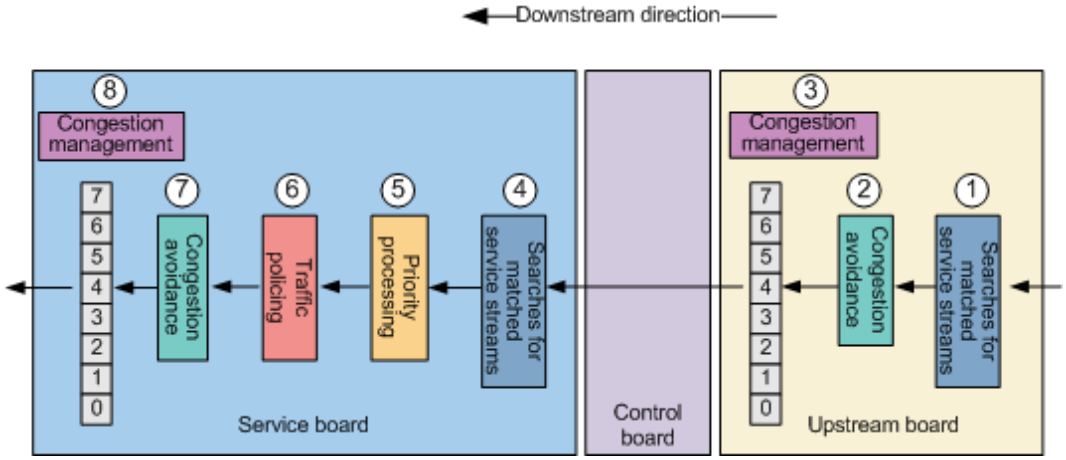
Step	QoS Policy	Example
6	Congestion avoidance: When congestion occurs during enqueuing on an inbound interface, the control board drops non-conformant packets to avoid further congestion.	To promptly drop packets based on priority, set the early drop threshold to 30% for 0-priority Internet services. After this setting takes effect, if 30% of the packets that arrive at a queue are Internet service packets, subsequent Internet service packets will be dropped.
7	Congestion management: When congestion occurs during dequeuing on an outbound interface, the control board uses queuing mechanisms to provide QoS guarantees to high-priority packets preferentially.	Use the SP scheduling algorithm to ensure that voice services with priority 5 are scheduled preferentially and Internet services with priority 0 are scheduled last when congestion occurs.

General QoS Processing - Downstream Direction

Figure 15-4 shows QoS processing in the downstream direction. Different from QoS processing in the upstream direction, QoS processing in the downstream direction is performed in the following way:

- Packets are first processed by the upstream board. After being switched on the control board, packets are processed by the service board.
- In the downstream direction, service flows are searched based on VLAN+MAC or SVLAN+CVLAN at the ingress of the upstream board.
- The service board does not perform traffic classification in the downstream direction. Instead, the service board searches for matched service flows based on a forwarding mode (VLAN+MAC or SVLAN+CVLAN).

Figure 15-4 QoS processing in the downstream direction



 Copyright © Huawei Technologies Co., Ltd.

15.5 Traffic Classification

Traffic classification differentiates services by packet classification according to the characteristics of user Ethernet packets and certain rules, to implement different processing operations and provide different services.

15.5.1 Introduction to Traffic Classification

Traffic

Service flow or data streams are a type of traffic. A service flow is a set of packets that have common properties. For example, the Internet service packets of a user are one service flow and voice service packets of the user are another service flow.

In the MA5800, a service flow is also called a service port.

For more details about the traffic, please refer to "**10.3 Service Flow**".

Traffic Classification

Traffic classification is a technique that categorizes user packets into different classes based on the properties of Ethernet packets and predefined QoS rules. Traffic classes are processed differently and this allows for the provision of differentiated services.

For example, to provide Internet, voice, and IPTV services simultaneously to the same user, the service packets must be separated into three classes.

Purpose

Traffic classification is used to support concurrent uses of multiple services. It differentiates service flows allowing predefined QoS guarantees to be provided for each of the service flows. The system performs service mappings based on service flows and performs QoS actions based on the service mappings. The QoS actions include switching between a user VLAN and the network VLAN, upstream and downstream CAR policing, priority marking, and queue scheduling.

15.5.2 Implementation Principle of Traffic Classification

Traffic Classification Process

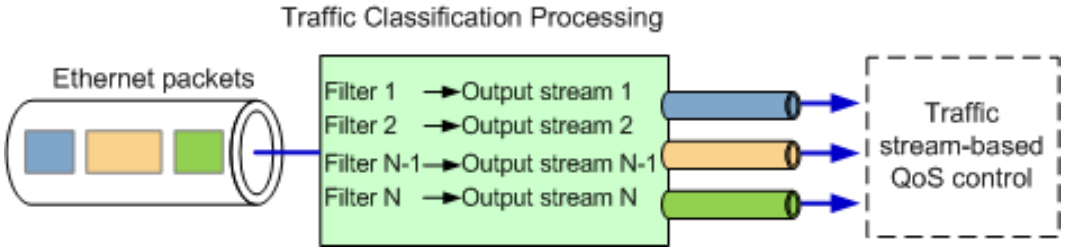
Figure 15-5 shows the traffic classification process. Traffic classification is performed using classifiers. Each classifier includes a filter (one or more classification rules as described in **Table 15-1**) and an output flow. The classifiers separate Ethernet packets received by a device into different traffic classes. QoS guarantees are then provided for each of the traffic classes.

Traffic classification is based on Ethernet packets. For PON access mode, the system segments and reassembles incoming service packets to recover Ethernet frames. The system then performs traffic classification for the Ethernet frames.

 Copyright © Huawei Technologies Co., Ltd.

Therefore, each GEM port can be regraded as a logical port of an Ethernet port. Traffic classification is performed based on this logical port and this logical port receives and sends Ethernet frames.

Figure 15-5 Traffic classification process



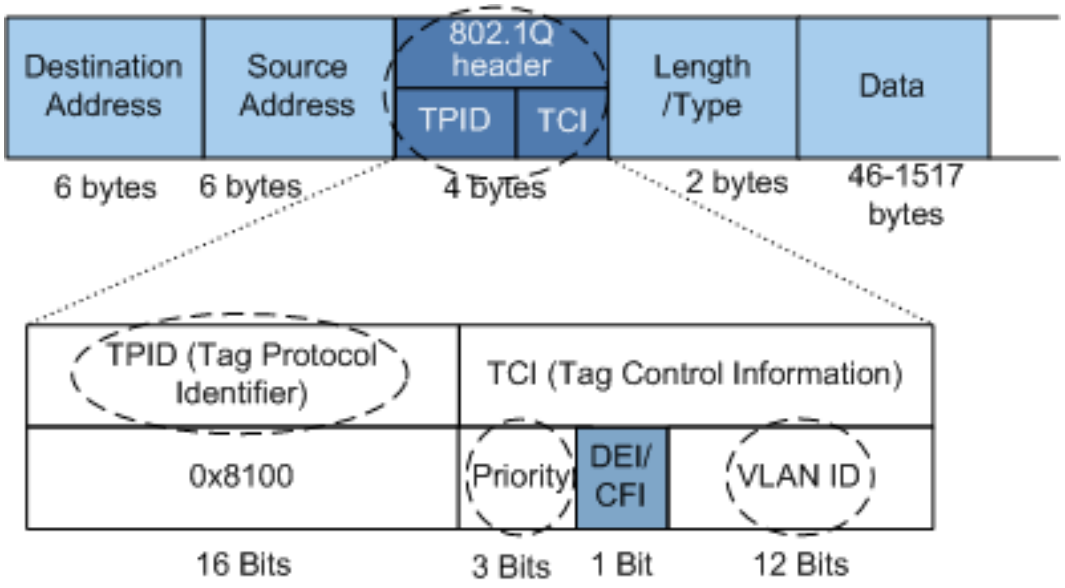
The system performs traffic classification only for upstream packets. Meanwhile, the system learns the MAC addresses of the packets when the packets are forwarded based on VLAN+MAC forwarding mode, and records the indexes of service flows. In the downstream direction, the system searches for the service flow based on the VLAN and MAC address.

Traffic Classification Method and Application

Traffic classification is performed to distinguish user services based on the properties of Ethernet packets. Traffic classification is usually performed based on the following fields in Ethernet frame headers (shown in the **Figure 15-6**):

- VLAN
- Priority
- Ethernet type

Figure 15-6 Ethernet frame format defined by 802.1q



 Copyright © Huawei Technologies Co., Ltd.

For detailed traffic classification rules, see **Table 15-1**.

Table 15-1 Traffic classification rules

Traffic Classification Rule	Application Scenario
Based on CVLAN	If services have different C-VLANs, traffic classification can be performed based on the C-VLANs. For example, traffic classification can be based on the C-VLAN of 100 for Internet services and 200 for voice services.
Based on priority	If the user-side priorities are different, traffic classification can be performed based on priority. For example, the priority of the Internet access service is 1 and the priority of the voice service is 5.
Based on Ethernet type	If the user-side Ethernet types are different, traffic classification can be performed based on Ethernet type. For example, traffic classification can be performed based on IPoE packets or PPPoE packets.
Based on C-VLAN +Ethernet type	If a single field cannot be used to distinguish the service type of packets, traffic classification can be based on combinations of the three fields in Ethernet packets. For example, if services have the same VLANs and priorities, the C-VLAN+Ethernet type combination can be used to distinguish the services.
Based on C-VLAN+priority	
Based on S-VLAN+C-VLAN	This rule is used when users of a service type must be identified uniquely. For example, if a C-VLAN is allocated to a specific user, the S-VLAN+C-VLAN combination can be used to identify this user.
Based on other-all	When a common service and a transparent LAN service exist in a system, other-all can be used to distinguish the two services. For example, a system has an Internet service (VLAN 100) and other transparent LAN services. If the Internet service must be distinguished from the other transparent LAN services, the transparent LAN services can be mapped into the other-all traffic class.

NOTE

- S-VLAN refers to service VLAN and is usually used to identify a service. In a service flow, S-VLAN is also called network-side VLAN or outer VLAN.
- C-VLAN refers to customer VLAN and is usually used to identify a user. In a service flow, C-VLAN is also called inner VLAN.

Generally, one logical port supports only one traffic classification rule, excluding the following two situations:

 Copyright © Huawei Technologies Co., Ltd.

- The single-tagged VLAN-based (**user-vlan** { **untagged** | *user-vlanid* }) traffic classification and the double-tagged VLAN-based (**double-vlan outer-vlan** *vlanid* **inner-vlan** *vlanid*) traffic classification can coexist.
- The single-service-based traffic classification and multi-service-based traffic classification can coexist on the port whose PVC encapsulation type is auto-sensing (**encapsulation type auto**). The system automatically matches the single-service flows or multi-service flows according to the learned PVC ATM adaptation layer type 5 (AAL5). Then, the single-service flows are used for IPoA/PPPoA services and multi-service flows are used for the Ethernet service.

Packet Matching Priority

- Packets are matched first to the rules that define specific classifiers (for example, C-VLAN or priority) and, if they cannot be matched to a specific traffic class, they are matched to the other-all class.
- If no other-all traffic class has been configured, all incoming packets are matched with specific traffic classes.
- If incoming packets cannot be matched with any traffic class, they are dropped.

15.5.3 Configuring the Traffic Classification

Context

A service flow is the result of traffic classification based on physical ports or logical ports.

Procedure

Step 1 Run the **service-port** command to perform traffic classification.

```
huawei(config)#service-port
{ desc<K>|index<U><0,131071><U>|remote-desc<K>|source<K>|statistics<K>|uplink-po
rt<K>|vlan<K> }: vlan
{ INTEGER<1-4093><U>|name<K> }: 100
{ eth<K>|gpon<K>|port<K>|vdsl<K> }: eth
{ STRING<5-18><S> }: 0/2/0
{ <cr>|bundle<K>|inbound<K>|multi-service<K>|rx-cttr<K>|tag-transform<K> }: mult
i-service
{ double-vlan<K>|user-8021p<K>|user-encap<K>|user-vlan<K> }: user-vlan
{ INTEGER<1-4095><U>|other-all<K>|priority-tagged<K>|untagged<K> }:
```

The key parameters related to traffic classification for this command are as follows:

- **multi-service**: indicates multiple services. Traffic classification is required when a service port carries multiple services.
- **double-vlan**: Traffic classification is performed based on S-VLAN+C-VLAN.
- **user-8021p**: Traffic classification is performed based on the user-side 802.1p priority.
- **user-encap**: Traffic classification is performed based on the user-side encapsulation type, IPoE or PPPoE.

 Copyright © Huawei Technologies Co., Ltd.

- **user-vlan**: Traffic classification is performed based on the user-side VLAN. Valid values for the user-side VLAN are 1-4095, untagged, priority-tagged (the user packet is tagged as VLAN 0), and other-all.

NOTE

For details on this command, see "Service Virtual Port Configuration" in Command Reference.

Step 2 Run the **display service-port** command to verify that the traffic classification has been applied to the service port.

----End

Example

- Traffic classification based on C-VLAN

  To perform traffic classification for Ethernet port 0/2/0 based on C-VLAN 100, do as follows:

  ```
  huawei(config)#service-port vlan 8 eth 0/2/0 multi-service user-vlan 100
  ```

- Traffic classification based on Ethernet type

  To perform traffic classification for Ethernet port 0/2/0 based on Ethernet type PPPoE, do as follows:

  ```
  huawei(config)#service-port vlan 8 eth 0/2/0 multi-service user-encap pppoe
  ```

- Traffic classification based on C-VLAN+Ethernet type

  To perform traffic classification for Ethernet port 0/2/0 based on C-VLAN 100 and Ethernet type PPPoE, do as follows:

  ```
  huawei(config)#service-port vlan 8 eth 0/2/0 multi-service user-vlan 100 user-encap pppoe
  ```

- Traffic classification based on C-VLAN+priority

  To perform traffic classification for Ethernet port 0/2/0 based on C-VLAN 100 and priority 3, do as follows:

  ```
  huawei(config)#service-port vlan 8 eth 0/2/0 multi-service user-vlan 100 user-8021p 3
  ```

- Traffic classification based on C-VLAN+S-VLAN

  To perform traffic classification for Ethernet port 0/2/0 based on S-VLAN 100 and C-VLAN 10, do as follows:

  ```
  huawei(config)#service-port vlan 8 eth 0/2/0 multi-service double-vlan outer-vlan 100 inner-vlan 10
  ```

15.6 Priority Marking

According to different priority marking policies, the inner and outer VLAN priorities can be set for service-ports, or the user-side priority can be copied for service-ports.

15.6.1 Introduction to Priority Marking

Definition

Priority processing is a process of marking or re-marking the priority for a packet so that equipment or network can process the packet based on the defined priority. This process is performed in the following way:

 Copyright © Huawei Technologies Co., Ltd.

- Priorities of packets are usually marked at an inbound interface of equipment or a network and is re-marked inside the equipment or network.
- Packet priorities include both forwarding and drop priorities. These two priorities form the basis for QoS processing. With respect to drop priority processing, the equipment or network uses the drop eligibility indicator (DEI) in an Ethernet packet to mark the color of the packet.

Purpose

Priority processing is the basis for equipment or a network to schedule packets. When congestion occurs, equipment or a network schedules packets based on priorities.

15.6.2 Basic Concepts of Priority Marking

Priorities that equipment processes mainly include the VLAN priority (802.1p priority) and IP precedence.

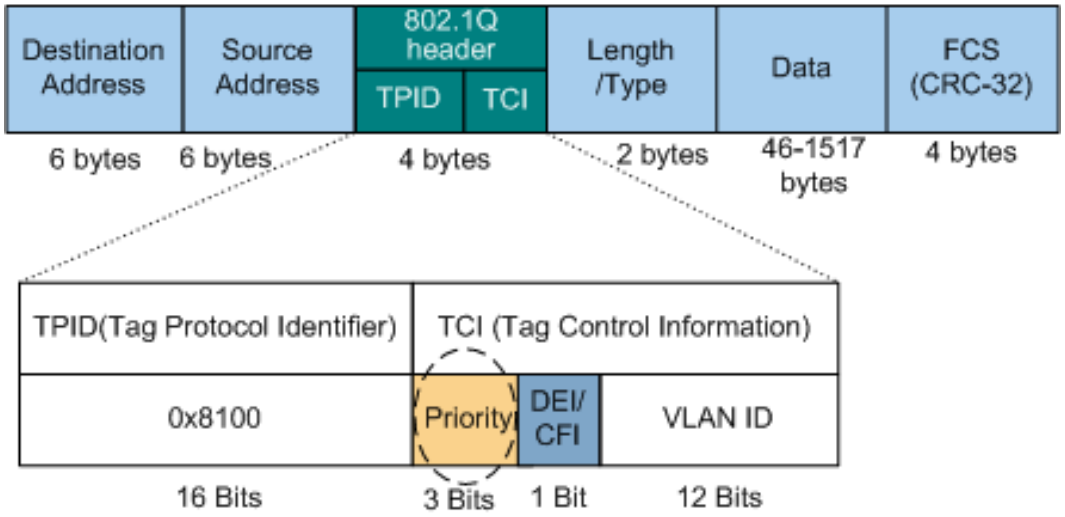
VLAN Priority

VLAN priority, also called 802.1p priority, refers to the packet priority defined at the link layer. This priority represents the class of service (CoS). As defined in IEEE 802.1q, the VLAN priority uses three bits in the VLAN tag (**Figure 15-7** shows the position of the VLAN priority in an Ethernet frame).

As shown in the figure, the Priority field indicates the 802.1p priority. This field consists of three bits. The value of the three bits ranges from 0 to 7. The value 0 is the lowest priority and 7 the highest. The priority values set in these fields determine the order in which packets are transmitted when congestion occurs on a port.

The DEI field consists of one bit and represents the drop eligible indicator defined in the 802.1ad protocol. It is used to color the packet. For example, the value 0 means green and 1 means yellow. When the function of is enabled, yellow packets are dropped with preference in case of congestion.

Figure 15-7 Ethernet frame format defined by 802.1q



 Copyright © Huawei Technologies Co., Ltd.

IP Precedence

The IP protocol defines differentiated services code point (DSCP) and type of service (ToS). They occupy the same field (one byte) in an IP header. IP bearer network devices schedule and forward packets based on the DSCP or ToS filled to provide QoS guarantees for different services.

ToS in the IP header specifies a traffic class for a packet rather than a priority (which is determined by the device). The ToS field consists of eight bits, including a 3-bit IP precedence sub-field, 4-bit ToS sub-field, and one reserved bit (set to 0). The four bits in the ToS sub-field represent the minimum latency, maximum throughput, maximum reliability, and minimum cost. Only one of the four bits can be set to 1. If all the four bits are set to 0, the service is a common service.

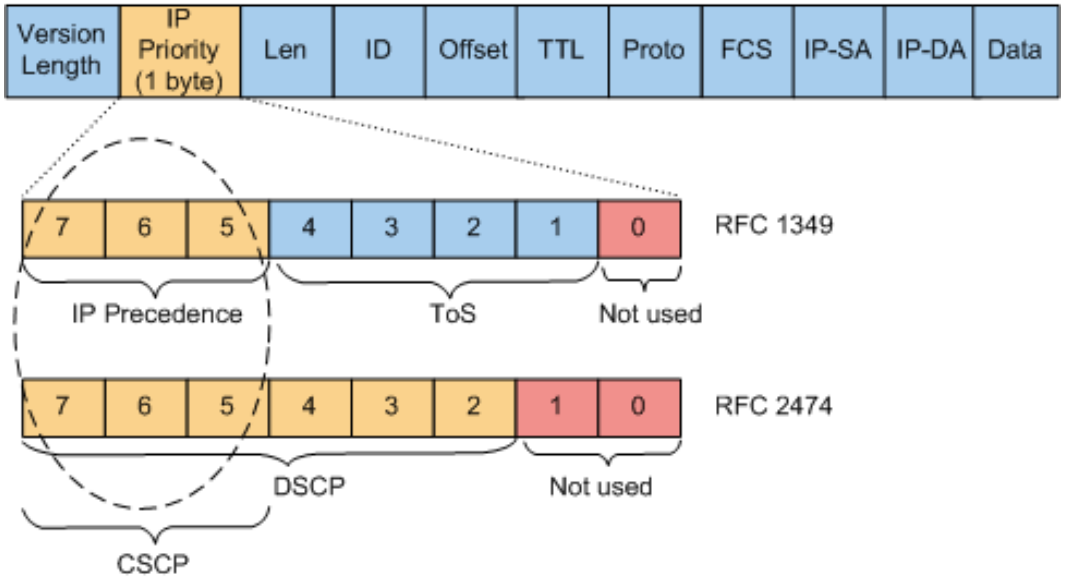
DSCP is defined in the RFC 2474, it is a re-defined object based on an IPv4 type of service (ToS) and an IPv6 traffic class.

NOTE

The traffic class (TC) field in the IPv6 packet header has the same functions as the IP priority field in the IPv4 packet header. This topic uses the IP priority field in the IPv4 packet header as an example to describe the IP priority.

As shown in **Figure 15-8**, the six most significant bits in the DS field (bits 7-2) are used as the DS CodePoint (DSCP) and the two least significant bits (bits 1 and 0) are reserved. The three most significant bits in the DS field (bits 7-6-5) are the class selector code point (CSCP), which indicates a DSCP type.

Figure 15-8 IPv4 packet format



DSCP is used to select per-hop behavior (PHB) on each node of a network. PHBs describe the visible external behavior when the DS node is used for data stream aggregation. IETF has defined the following types of PHBs:

 Copyright © Huawei Technologies Co., Ltd.

- Class selector (CS)
- Expedited forwarding (EF)
- Assured forwarding (AF)
- Best-effort (BE)

Table 15-2 shows the structure of PHBs.

Table 15-2 Structure of PHBs

PHB	Bit 7-6-5	Bit 4-3-2
CS	aaa (Remark)	000
BE	000	000
EF	101	110
AF	bbb (Remark)	cc0 (Remark)

Remark: a, b and c indicates a single bit, whose value is 0 or 1. Where,

- The "aaa" has eight values, from 000 to 111, it corresponds to the decimal number 0-7. The 3 bits can map with IP precedence individually.
- The "bbb" has four values: 001, 010,011 and 100, it corresponds to the decimal number 1-4.
- The "cc" has three values, from 01 to 11, it corresponds to the decimal number 1-3.

Table 15-3 shows the common DSCP service types and corresponding priorities.

Table 15-3 Common DSCP service types and corresponding priorities

DSCP type	IPv4 precedence/ 802.1P priority	DSCP value (binary)	Application
BE	0	0	Internet
AF1	1	001 010	Leased Line
AF1	1	001 100	Leased Line
AF1	1	001 110	Leased Line
AF2	2	010 010	IPTV VoD
AF2	2	010 100	IPTV VoD
AF2	2	010 110	IPTV VoD
AF3	3	011 010	IPTV Broadcast
AF3	3	011 100	IPTV Broadcast
AF3	3	011 110	IPTV Broadcast

 Copyright © Huawei Technologies Co., Ltd.

DSCP type	IPv4 precedence/ 802.1P priority	DSCP value (binary)	Application
AF4	4	100 010	NGN/3G Signaling
AF4	4	100 100	NGN/3G Signaling
AF4	4	100 110	NGN/3G Signaling
EF	5	101 110	NGN/3G voice
CS6	6	110 000	Protocol
CS7	7	111 000	Protocol

15.6.3 Priority Sources

Priority processing includes copying, designating, and mapping inner and outer VLAN priorities.

Priority Processing for an Outer VLAN (or Single-Tagged VLAN) of Ethernet Service Flows

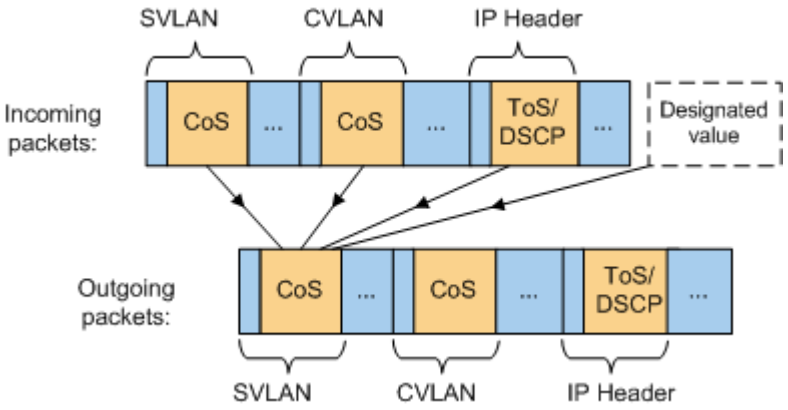
Figure 15-9 shows how the priority of an outer VLAN or the single-tagged VLAN is processed. The priority of an outer VLAN can be derived from multiple sources:

- Copied from the outer VLAN priority (user-cos) of an incoming packet
- Copied from the inner VLAN priority (user-inner-cos) of an incoming packet
- Copied or mapped from the IP ToS priority (user-tos) of an incoming packet
- Mapped from the IP DSCP priority (user-dscp) of an incoming packet
- Copied from the priority of a designated packet (prival)

NOTE

Currently, the downstream equipment of a service flow cannot copy the user-dscp priority.

Figure 15-9 Priority processing for an outer VLAN (or single-tagged VLAN) of Ethernet Service Flows



 Copyright © Huawei Technologies Co., Ltd.

Some boards support several priority mapping choices:

- 802.1p priority -> 802.1p priority (To implement this mapping, run the **pbits-to-pbits mapping table** command. It's only valid for inner VLAN priority)
- IP priority -> 802.1p priority (To implement this mapping, run the **ipprec-to-pbits mapping table** command.)
- DSCP priority -> 802.1p priority (To implement this mapping, run the **dscp-to-pbits mapping table** command.)

Figure 15-10 shows the mappings between incoming and outgoing priorities of packets.

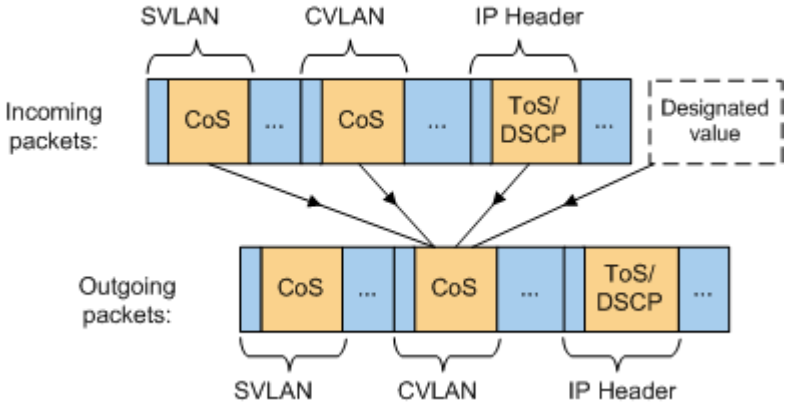
Figure 15-10 Mapping between the incoming and outgoing priorities of packets

802.1p/IP priority (Incoming packets)	DSCP priority (Incoming packets)	802.1p priority of outgoing packets (Copy priority by default, and priority can't be modified)	802.1p priority of outgoing packets (Mapping, priority can be modified)
0	0-7	0	0
1	8-15	1	1
2	16-23	2	2
3	24-31	3	3
4	32-39	4	3
5	40-47	5	3
6	48-55	6	4
7	56-63	7	5

Priority Processing for an Inner VLAN of Ethernet Service Flows

Figure 15-11 shows how the priority of an inner VLAN is processed. The sources of an inner VLAN priority are the same as those of an outer VLAN priority.

Figure 15-11 Priority processing for an inner VLAN



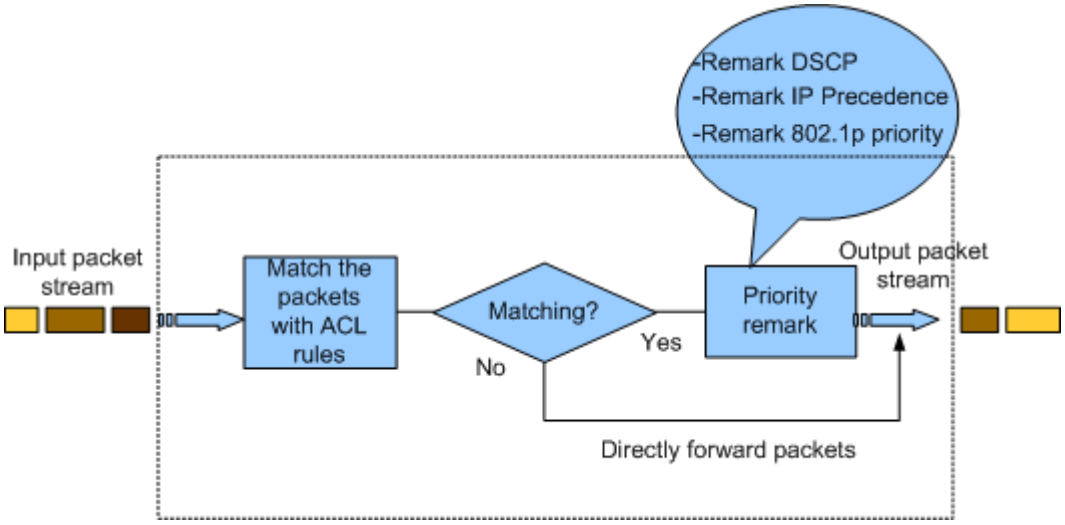
 Copyright © Huawei Technologies Co., Ltd.

Priority Processing for ACL-based Packets

Figure 15-12 shows the policy of priority processing for ACL-based packets. The system processes the input packets according to ACL rules. After packets match ACL rules, the system remarks these packets by running the **traffic-priority** command.

- Specifies the DSCP priority of the packets.
- Specifies or copies the 802.1p priority of packets as the IP precedence priority of packets.
- Specifies or copies the IP precedence priority of packets as the 802.1p priority of packets.

Figure 15-12 Priority processing for ACL-based packets



15.6.4 Implementation Principle of Priority Marking

Priority processing applies to common data packets and specific protocol packets. The priority processing mechanisms for data packets and protocol packets are different, because protocol packets are captured and processed by the CPU.

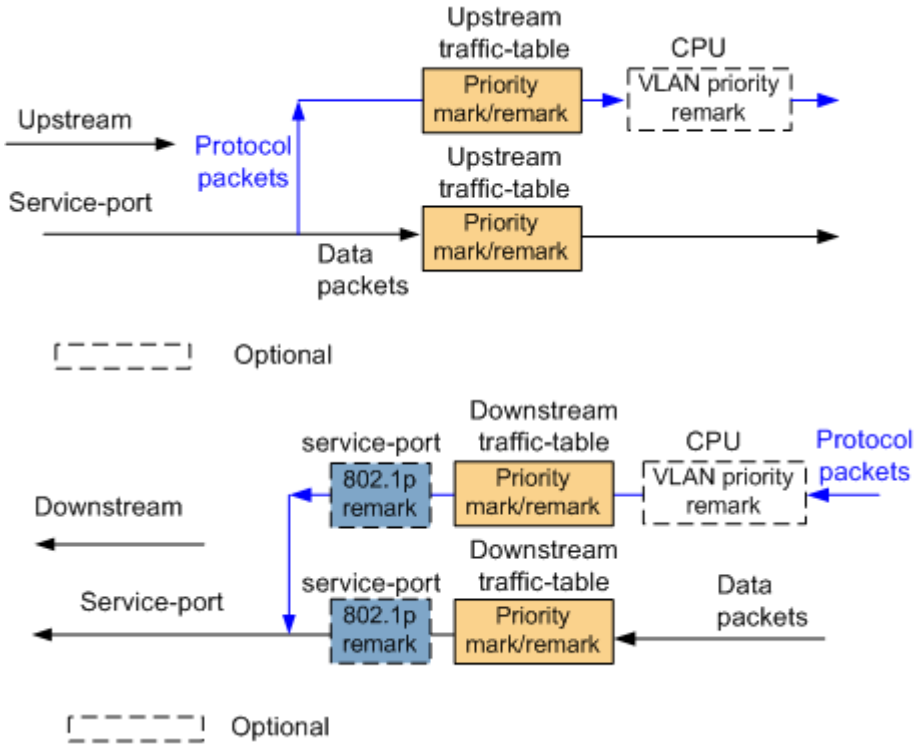
In addition, priority processing is affected by the following factors:

- Attribute parameters of service flows (whether user-802.1p is configured)
- **Priority source**
- VLAN priority
- Packet forwarding mode (Layer 2 or Layer 3 forwarding)

Figure 15-13 describes packet priority processing in the non-service-bundle scenario.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-13 Packet priority processing in the non-service-bundle scenario



Note that the packet priority processing described in the preceding scenarios is a general processing model. Not all types of packets strictly comply with this model. For details on priority processing for packets of a particular type, refer to the priority processing mechanism for the packet type.

15.6.4.1 Priority Processing for Data Packets

Priority Processing for Data Packets in the Non-Service-Bundle Scenario

Priority processing for data packets in the non-service-bundle scenario has the following characteristics:

- The service flow priority is irrelevant to the VLAN priority (Run the **vlan priority** *vlanid priority* command to set VLAN priority).
- Priority processing in the upstream /downstream direction is based on the priority processing policy (specify or copy/map priorities, as shown in **Figure 15-14** and **Figure 15-15**) configured in the upstream /downstream traffic profile.
- If the user-cos, user-tos, or user-inner-cos parameter is selected but no such priority exists in the packet (in other words, priority copying fails), the system uses the default 802.1p priority for the packet.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-14 Priority processing for upstream data packets in the non-service-bundle scenario

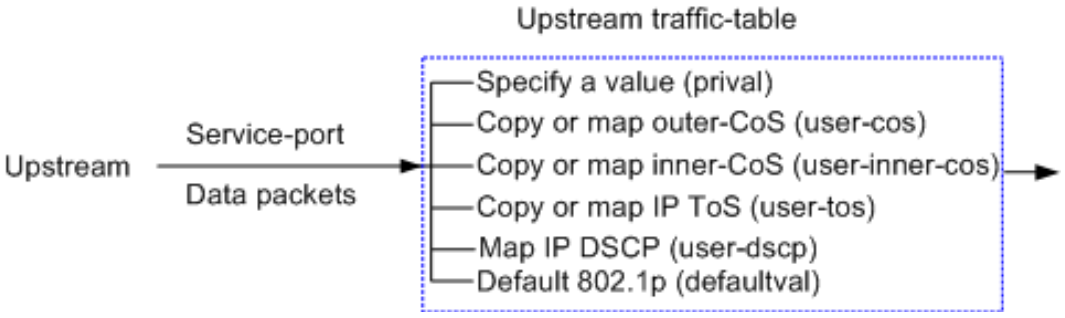


Figure 15-15 Priority processing for downstream data packets in the non-service-bundle scenario

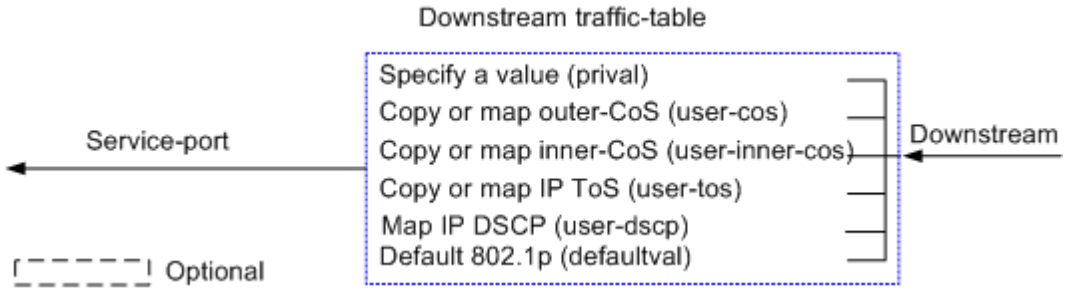


Figure 15-16 shows the detailed processing results.

Figure 15-16 Priority processing results of upstream and downstream data packets in the non-service-bundle scenario

service-port		traffic table				vlan priority	Forwarding mode	ONT gemport mapping	xPON & P2P System Behavior	
user-8021p	user-vlan	priority	user-cos	user-tos	deafult				Upstream direction	Downstream direction
-	√	√	-	-	-	NA	L2	Priority Mapping	traffic table priority	traffic table priority
-	√	√	-	-	-	NA	L3	Priority Mapping	traffic table priority	traffic table priority
√	√	√	-	-		NA	L2	Priority Mapping	traffic table priority	traffic table priority
√	√	√	-	-		NA	L3	Priority Mapping	traffic table priority	traffic table priority
-	√	-	√	-	√	NA	L2	Priority Mapping	user-cos	user-cos
-	√	-	√	-	√	NA	L3	Priority Mapping	user-cos	user-cos
√	√	-	√	-	√	NA	L2	Priority Mapping	user-cos	user-cos
√	√	-	√	-	√	NA	L3	Priority Mapping	user-cos	user-cos
-	√	-	-	√	√	NA	L2	Priority Mapping	user-tos	user-tos
-	√	-	-	√	√	NA	L3	Priority Mapping	user-tos	user-tos
√	√	-	-	√	√	NA	L2	Priority Mapping	user-tos	user-tos
√	√	-	-	√	√	NA	L3	Priority Mapping	user-tos	user-tos
√: Configured		-:Unconfigured								

15.6.4.2 Priority Processing for DHCP & PPPoE Packets

 Copyright © Huawei Technologies Co., Ltd.

Priority Processing for DHCP & PPPoE Packets in the Non-Service-Bundle Scenario

DHCP protocol packets can work in Layer 2 or Layer 3 relay mode. The DHCP protocol packets in this document refer to DHCP Option82 packets in Layer 2 relay mode and DHCP Layer 3 relay packets.

PPPoE protocol packets work only in Layer 2 mode. The PPPoE packets in this document refer to PPPoE PITP packets.

When DHCP and PPPoE packets work in Layer 2 mode, the priority processing mechanism is the same as that for common data packets, which has the following characteristics:

- The service flow priority is irrelevant to the VLAN priority.
- Priority processing in the upstream /downstream direction is based on the priority processing policy (specify or copy/map priorities) configured in the upstream /downstream traffic profile.
- If the **user-cos**, **user-tos**, or **user-inner-cos** parameter is selected but no such priority exists in the packet (in other words, priority copying fails), the system uses the default 802.1p priority for the packet.

  NOTE

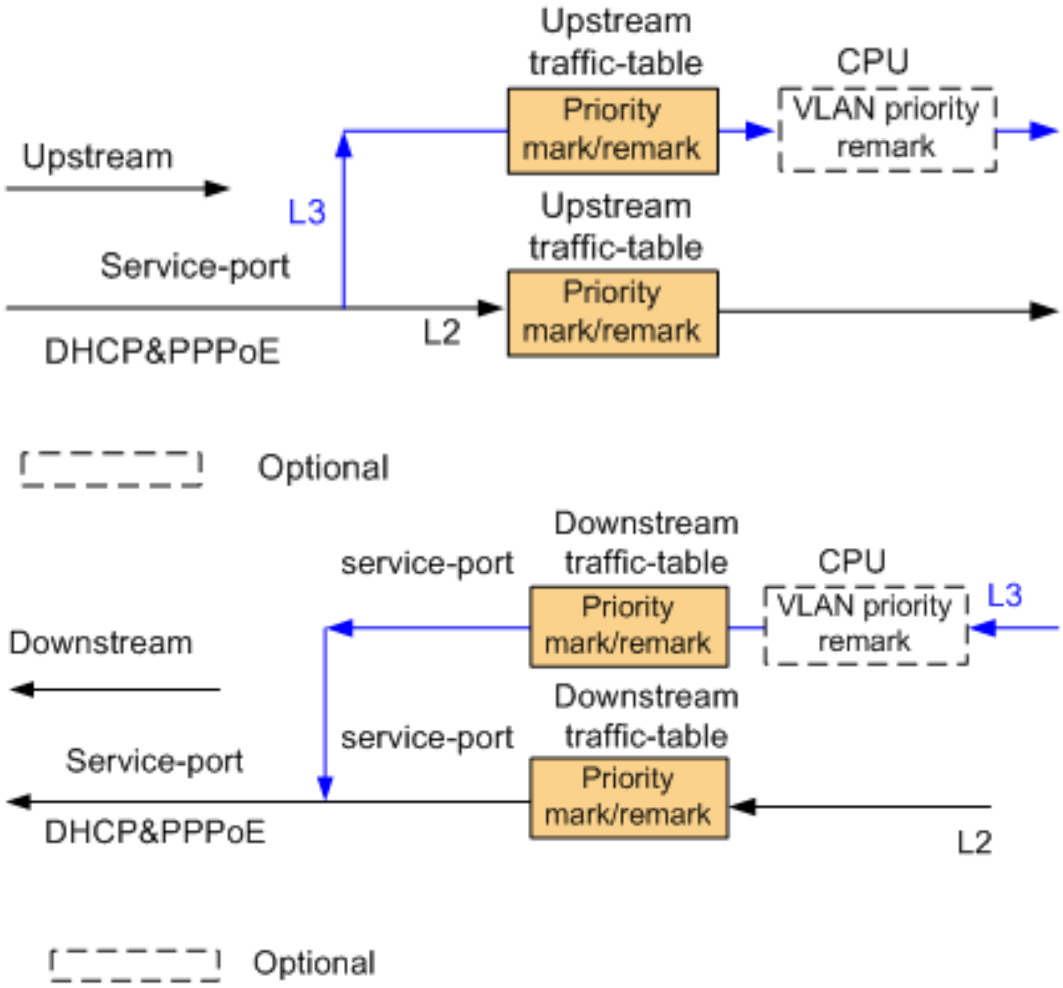
  When the user-tos priority policy is specified in the traffic profile, ToS copying fails for PPPoE PITP packets. When this occurs, the default 802.1p priority in the traffic profile will be used.

When DHCP packets work in Layer 3 mode, the priority processing mechanism has the following characteristics:

- Priority processing in the upstream direction is based on the priority processing policy (specify or copy/map priorities) configured in the upstream traffic profile.
- When the **prival** parameter or the **user-cos**/**user-inner-cos**/**user-tos** parameter in the upstream traffic profile is set to 0, the network-side VLAN priority takes effect.
- In the downstream direction, if the network-side priority of the packet is **0**, the packet priority is remarked as the user-side VLAN priority, and then is remarked based on the downstream traffic profile.
- In the downstream direction, if the priority processing policy specified in the traffic profile is **user-tos**, the user-side priority of the packet is fixedly remarked as **0**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-17 Priority processing for upstream and downstream DHCP & PPPoE packets



Priority Processing for DHCP & PPPoE Packets in the Service-Bundle Scenario

Priority processing for DHCP & PPPoE packets in the service-bundle scenario has the following characteristics:

- The priority of upstream packets is the priority specified for the service-bundle group. The settings in the traffic profile do not apply.
- The priority processing for downstream packets is the same as that in the non-service-bundle scenario.

15.6.4.3 Priority Processing for DHCPv6 Packets

Priority Processing for DHCPv6 Packets in the Non-Service-Bundle Scenario

DHCPv6 protocol packets can work in Layer 2 or Layer 3 relay mode. The DHCPv6 protocol packets in this document refer to DHCP Option18/Option37 packets in Layer 2 relay mode and DHCP Layer 3 relay packets.

 Copyright © Huawei Technologies Co., Ltd.

When DHCPv6 packets work in Layer 2 mode, the priority processing mechanism is the same as that for DHCPv4 packets in Layer 2 mode, which has the following characteristics:

- The service flow priority is irrelevant to the VLAN priority.
- Priority processing in the upstream /downstream direction is based on the priority processing policy (specify or copy/map priorities) configured in the upstream /downstream traffic profile.
- If the **user-cos**, **user-tos**, or **user-inner-cos** parameter is selected but no such priority exists in the packet (in other words, priority copying fails), the system uses the default 802.1p priority for the packet.

When DHCPv6 packets work in Layer 3 mode, the priority processing mechanism is the same as that for DHCPv4 packets in Layer 3 mode, which has the following characteristics:

- Priority processing in the upstream direction is based on the priority processing policy (specify or copy/map priorities) configured in the upstream traffic profile.
- When the **prival** parameter or the **user-cos**/**user-inner-cos**/**user-tos** parameter in the upstream traffic profile is set to 0, the network-side VLAN priority takes effect.
- In the downstream direction, if the network-side priority of the packet is **0**, the packet priority is remarked as the user-side VLAN priority, and then is remarked based on the downstream traffic profile.
- In the downstream direction, if the priority processing policy specified in the traffic profile is **user-tos**, the user-side priority of the packet is fixedly remarked as **0**.

Priority Processing for DHCPv6 Packets in the Service-Bundle Scenario

Priority processing for DHCPv6 packets in the service-bundle scenario has the following characteristics:

- The priority of upstream packets is the priority specified for the service-bundle group. The settings in the traffic profile do not apply.
- The priority processing for downstream packets is the same as that in the non-service-bundle scenario.

15.6.4.4 Priority Processing for IGMP Packets

The IGMP packets in this document refer to IGMP packets for IPv6 and IPv4. The priority processing applies to IGMP packets for both IPv4 and IPv6 unless otherwise specified.

Priority Processing for IGMP Packets in the Non-Service-Bundle Scenario

The IGMP mode for multicast VLANs can be IGMP snooping or IGMP proxy. The priority processing mechanism for IGMP packets varies with the IGMP mode.

Figure 15-18 describes the priority processing mechanism for IGMP packets in IGMP snooping mode.

 Copyright © Huawei Technologies Co., Ltd.

- Priority processing in the upstream direction is based on the priority processing policy (specify or copy/map priorities) configured in the upstream traffic profile.
- In the downstream direction, if the multicast VLAN tag mode is set to "dual" by running the **igmp multicast-tag dual** command, then the inner VLAN priority of IGMP packet is the same as the outer VLAN priority.
- In the downstream direction, priority processing is implemented based on the settings in the downstream traffic profile.

  NOTE

  The current version does not support the copying policy for IGMP protocol packets in the downstream direction. Therefore, when the priority processing policy specified is user-cos or user-tos, the default priority (defaultval) in the downstream traffic profile is used.

Figure 15-18 Priority processing for IGMP packets in IGMP snooping mode

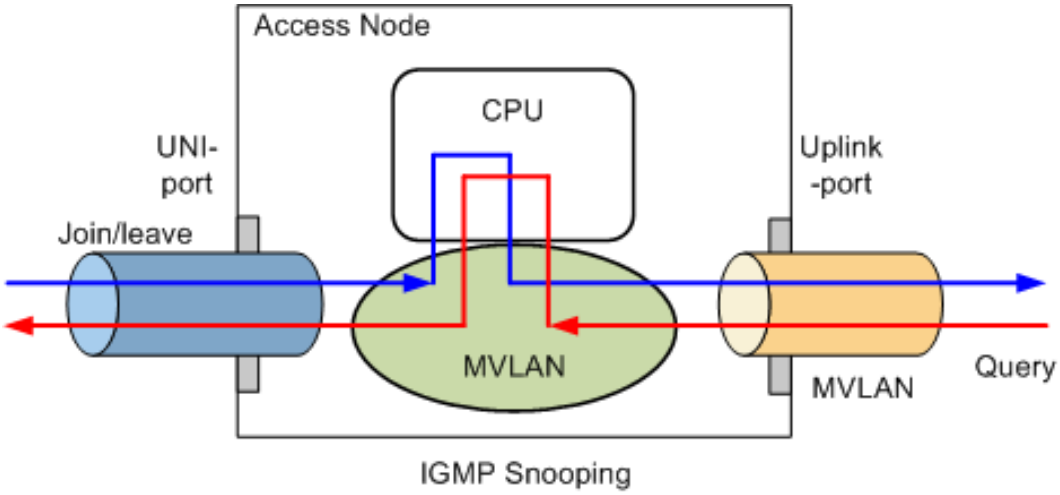


Figure 15-19 describes the priority processing mechanism for IGMP packets in IGMP proxy mode.

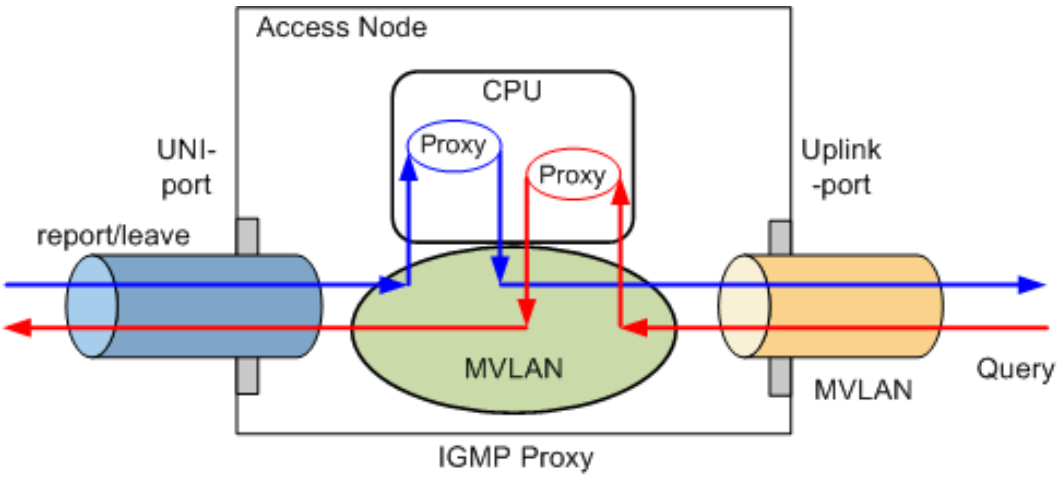
- The priority processing in the upstream direction depends on the specified IGMP packet priority. Command for setting the IGMP packet priority: **igmp priority** (IPv4)/**igmp ipv6 priority** (IPv6)
- In the downstream direction, priority processing is implemented based on the settings in the downstream traffic profile.

  NOTE

  The current version does not support the copying policy for IGMP protocol packets in the downstream direction. Therefore, when the priority processing policy specified is user-cos or user-tos, the default priority (defaultval) in the downstream traffic profile is used.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-19 Priority processing for IGMP packets in IGMP proxy mode



Priority Processing for IGMP Packets in the Service-Bundle Scenario

Priority processing for IGMP packets in the service-bundle scenario has the following characteristics:

- In IGMP proxy mode, the priority processing for upstream and downstream packets is the same as that in the non-service-bundle scenario.
- In IGMP snooping mode, the priority processing mechanism has the following characteristics:
  - The priority of upstream packets is the priority specified for the service-bundle group. The settings in the traffic profile do not apply.
  - The priority processing for downstream packets is the same as that in the non-service-bundle scenario.

15.6.4.5 Priority Processing for ARP Packets

ARP packets in this document refer to ARP packets sent by access nodes in the upstream and downstream directions. For ARP packets sent by non-access nodes, the priority processing mechanism is the same as that for common data packets.

For ARP packets, the upstream/downstream direction has a meaning that differs from the meaning for other protocol packets:

- Upstream direction: In this direction, an access node gives ARP responses or initiates ARP requests to the network side.
- Downstream direction: In this direction, an access node gives ARP responses or initiates ARP requests to the user side.

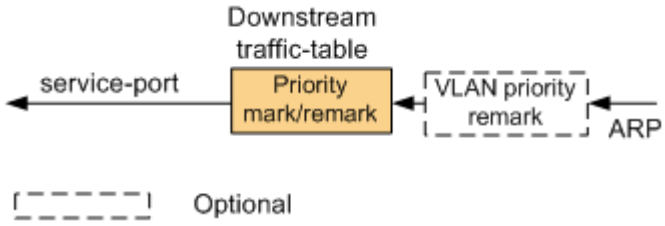
Priority Processing for ARP Packets in the Non-Service-Bundle Scenario

The priority processing mechanism for ARP packets has the following characteristics:

 Copyright © Huawei Technologies Co., Ltd.

- In the upstream direction, priority processing depends only on the VLAN priority and is irrelevant to the priority configuration of the service flow.
  - If no VLAN priority is configured, the packet priority will be remarked as 0 (the default priority for ARP packets).
  - If the VLAN priority is configured, the packet priority will be remarked as the VLAN priority.
- **Figure 15-20** describes the priority processing in the downstream direction.
  - The priority processing in the downstream direction involves VLAN priority remarking, priority remarking based on the traffic profile, and service flow 802.1p remarking.
  - If the service flow is created on the GPON/ETH board and the priority processing policy specified in the traffic profile is **user-cos**, the 802.1p remarking is skipped. In this case, if the VLAN priority is configured, the packet priority will be remarked based on the VLAN priority; if no VLAN priority is configured, the packet priority will be remarked as 0.
  - If the priority processing policy specified in the traffic profile is user-tos or user-inner-cos, priority copying fails because ARP packets do not have such fields. In this case, the packet priority is remarked as the default value (defaultval) specified in the traffic profile.

Figure 15-20 Priority processing for downstream ARP packets in the non-service-bundle scenario



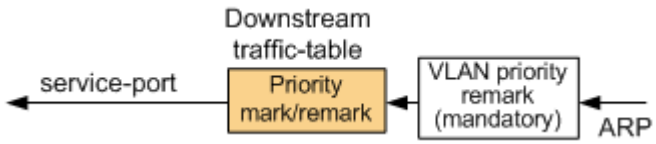
Priority Processing for ARP Packets in the Service-Bundle Scenario

In the upstream direction, priority processing for ARP packets is irrelevant to the priority configuration of the service flow and is the same as that in the non-service-bundle scenario.

In the service-bundle scenario, the priority of ARP packets sent by an access node to the user side must be the bundle-cos of the service flow; otherwise, the ARP packets fail to be received. Therefore, the VLAN priority must be configured.

Figure 15-21 describes the priority processing mechanism for downstream ARP packets in the service-bundle scenario. According to **Figure 15-21**, the priority processing is almost the same as that in the non-service-bundle scenario. The difference is that the VLAN priority is mandatory in this scenario while it is optional in the non-service-bundle scenario.

Figure 15-21 Priority processing for downstream ARP packets in the service-bundle scenario



 Copyright © Huawei Technologies Co., Ltd.

15.6.4.6 Priority Processing for ND Packets

This topic describes the priority processing for neighbor discovery (ND) packets. For detailed processing rules and results, refer to Priority Processing Rules.

ND packets apply to the IPv6 scenario. Similar to the ARP function for IPv4 packets, the ND function is used to determine the link-layer address and reachability of a neighboring node for IPv6 packet forwarding.

Because the function of ND packets is similar to that of ARP packets, the priority processing mechanism for ND packets is also similar to that for ARP packets. For details, refer to **Priority Processing for ARP Packets**. The difference is that

- **The default priority for ARP packets is 0 while that for ND packets is 6.**
- **The ARP packet does not carry any IP header, and therefore the user-tos copying in the downstream direction is not supported. In this case, the priority of the ARP packet is remarked as the default priority specified in the traffic profile. The ND packet carries an IP header, and therefore the user-tos copying in the downstream direction is supported. In this case, the priority of the ND packet is remarked as 6.**

15.6.5 Configuring the Priority Processing

Context

Priority processing is the basis for equipment or a network to schedule packets. When congestion occurs, equipment or a network schedules packets based on priority.

A priority processing policy is configured in a traffic profile for equipment.

Procedure

Step 1 Run the **traffic table ip** command to configure priority processing.

```
huawei(config)#traffic table ip index 11
cir<K>|name<K> }:cir
{ cir<U><64,10240000>|off<K> }:off
{ color-policy<K>|priority<K> }:priority
{ prival<U><0,7>|user-cos<K>|user-inner-cos<K>|user-tos<K> }:user-cos
{ defaultval<U><0,7>|mapping-profile<K> }:3
{ inner-priority<K>|priority-policy<K> }:inner-priority
{ inner-prival<U><0,7>|user-cos<K>|user-inner-cos<K>|user-tos<K> }:user-inner-cos
{ defaultval<U><0,7>|mapping-profile<K> }:3
{ priority-policy<K> }:priority-policy
{ priority-policy<E><Local-Setting,Tag-In-Package,Tag-In-Ingress-Package> }:tag-In-package
  Command:
        traffic table ip index 11 cir off priority user-cos 3 inner-priority u
ser-inner-cos 3 priority-policy tag-In-package

  Create traffic descriptor record successfully
  -------------------------------------------------------
  Traffic Table Index          : 11
  Traffic Table Name           : ip-traffic-table_11
  Specified Outer-Priority     : -
  Outer-Priority Mapping Source: user-cos
  Outer-Priority Mapping Index : 0
  Default Outer-Priority       : 3
  Specified Inner-Priority     : -
```

 Copyright © Huawei Technologies Co., Ltd.

```
Inner-Priority Mapping Source: user-inner-cos
Inner-Priority Mapping Index : 0
Default Inner-Priority       : 3
Specified DSCP               : -
DSCP Mapping Source          : -
DSCP Mapping Index           : -
CIR                        : off
CBS                        : off
PIR                        : off
PBS                        : off
Fix                        : off
CAR Threshold Profile        : -
Color Mode                 : color-blind
Coupling Flag              : enable
Rate Profile Index         : -
Rate Profile Name          : -
Color Policy               : dei
Color DEI                  : unmark-dei
Color Source               : outer-dei
Drop Precedence            : remarked-outer-priority
Priority Policy            : tag-pri
Que-Policy Profile ID        : -
Pri-Group Profile ID         : -
Pri-Group Profile Name       : -
Referenced Status          : not used
E2E Mapping Priority         : -
E2E Mapping Index            : -
E2E Default Priority         : -
Yellow WRR Weight            : auto
Green WRR Weight             : auto
-------------------------------------------------------------
```

The parameters for this command are as follows:

- **priority**: specifies the S-VLAN priority policy, including the priority source and queue scheduling policy. The upstream priority determines which queue that upstream packets enter. Valid values for this parameter are as follows:
  - ***prival***: specifies a priority for upstream and downstream packets. When **priority-policy** is set to **Local-Setting**, the priority is determined by the value of **Local-Setting**.
  - **user-cos**: copies the priority from the outer 802.1q tag of an incoming packet as the S-VLAN priority.
  - **user-inner-cos**: copies the priority from the inner 802.1q tag of an incoming packet as the S-VLAN priority.
  - **user-tos**: copies the priority from the IP ToS field of an incoming packet as the S-VLAN priority.
- **inner-priority**: specifies the C-VLAN priority policy. The parameter values are the same as those for the parameter that specifies the S-VLAN priority policy. C-VLAN and S-VLAN can be set independently.
- **priority-policy**: specifies the priority policy for queue scheduling. The queue scheduling priority of a packet is generally the same as the priority of the packet. If queue scheduling and packet priorities are different, the queue scheduling priority must be manually specified. Valid values for this parameter are as follows:
  - **Local-Setting**: uses a manually specified priority as the queue scheduling priority.
  - **Tag-In-Package**: uses the priorities of outgoing packets (after VLAN translation) as the queue scheduling priority.

 Copyright © Huawei Technologies Co., Ltd.

– **Tag-In-Ingress-Package**: uses the priorities of incoming packets (before VLAN translation) as the queue scheduling priority for the downstream direction.

NOTE

For details on this command, see "QoS Configuration" in "QoS Command" of Command Reference.

Step 2 Run the **display traffic table ip** command to verify the configuration result.

----End

Follow-up Procedure

After configuring a traffic profile, run the **service-port** command to bind service flows to the profile to ensure that different priority processing policies are applied to different types of traffic.

15.7 Traffic Policing

Before service providers provide subscribers with specific services, a service level agreement (SLA) is generally assigned, in which all service parameters are defined. To ensure that the user traffic can meet the SLA, the user traffic must be policed.

15.7.1 Introduction to Traffic Policing

Definition

Traffic policing (also called traffic policy) sets limits on the rate of incoming and outgoing service flow packets. Policing measures packet rates to monitor service flow bursts. Incoming and outgoing packets must meet user-defined conditions and, if they do not meet these conditions, the device that the packets pass through is configured to take countermeasures. For example, if the incoming packet rate exceeds a threshold, the device can take actions such as dropping or coloring the packets (re-setting the packet priorities).

Traffic policing is usually implemented using the committed access rate (CAR). In a PON system, upstream bandwidth conflict between ONUs is resolved by the dynamic bandwidth assignment (DBA) technology.

Purpose

Traffic policing enables carriers to achieve the following goals:

- Ensure that user traffic stays within the service level agreement (SLA).
- Ensure service quality by regulating outgoing traffic and suppressing burst traffic.
- Control the rate of broadcast packets using packet suppression.

15.7.2 Basic Concepts of Traffic Policing

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-22 shows the basic concepts of traffic policing and **Table 15-4** describes these concepts.

Figure 15-22 Key concepts of traffic policing

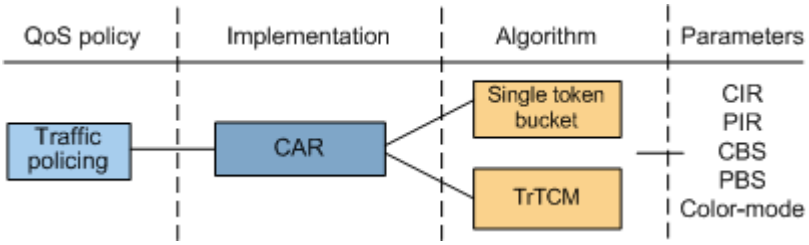


Table 15-4 Description of the basic concepts

Concept	Description
CAR	Stands for committed access rate. It is a technique commonly used to provide the specified rate for specific traffic. And it is widely used to limit the rate of ethernet ports, xDSL ports and xPON ports. CAR is usually implemented using token bucket algorithms.
Token bucket	A token bucket is a container that stores tokens, and it is used for controlling data traffic. A token bucket allows burst data transmission while controlling the traffic. There are two token bucket algorithms: single token bucket algorithm and dual token bucket algorithm. According to the principle of token bucket algorithms, a packet is transmitted when there is equivalent size of tokens in the token bucket. After a packet is transmitted, the number of tokens in the buckets decreases accordingly.
trTCM	Stands for two rate three color marker. It is defined in RFC2698 developed by the Internet Engineering Task Force (IETF). The trTCM algorithm sets the DEI bit for an ethernet packet based on two rates (PIR and CIR) and the burst size to mark the packet green, yellow, or red.
DEI	Stands for drop eligible indicator and is defined by 802.1ad. The DEI field in an ethernet packet consists of 1 bit and is the same as the CFI field defined by 802.1q. This field is used to mark an Ethernet packet with a corresponding color.
CIR	Stands for committed information rate. The unit is bit/s.
PIR	Stands for peak information rate. It provides users the maximum bandwidth when the system is idle. The unit is bit/s.

 Copyright © Huawei Technologies Co., Ltd.

Concept	Description
CBS	Stands for committed burst size. It is used to define the capacity of token bucket C, that is, the maximum burst IP packet size when information is transferred at the committed information rate. The unit is byte.
PBS	Stands for peak burst size. It is used to define the capacity of token bucket P, that is, the maximum burst IP packet size when the information is transferred at the peak information rate. The unit is byte.
Color mode	The trTCM algorithm uses two modes: color-blind and color-aware. In the color-blind mode, the trTCM algorithm assumes that the packet is uncolored. In the color-aware mode, the trTCM algorithm assumes that the packet is pre-colored.

15.7.3 Implementation Principle: CAR

Principle of Single Token Bucket

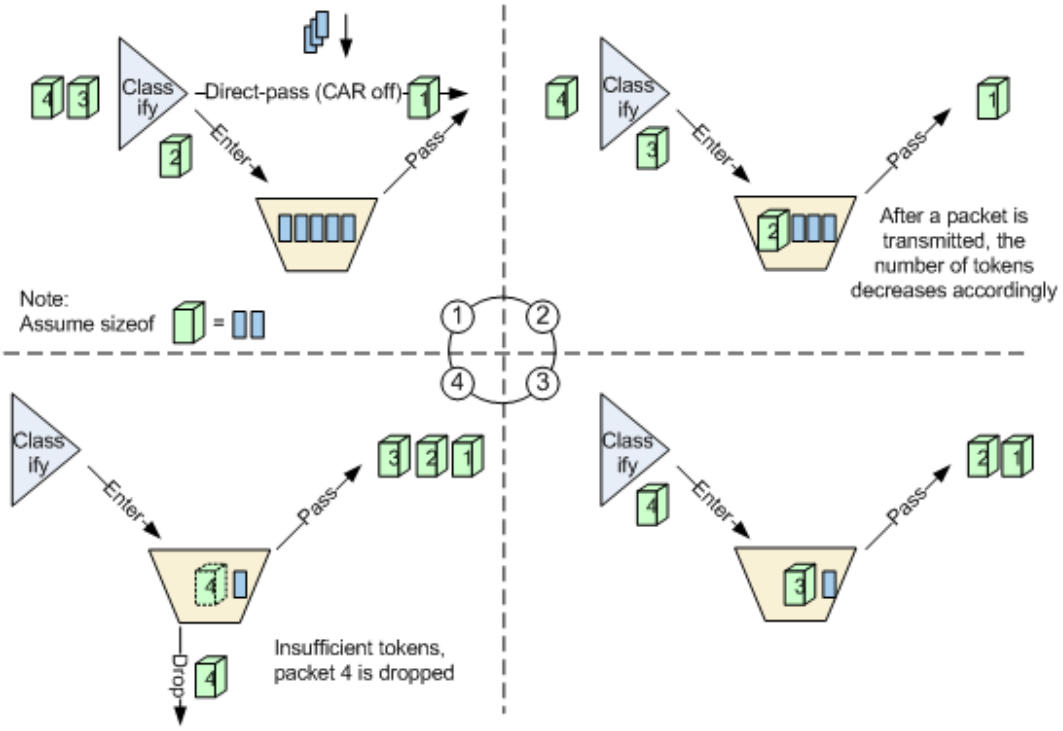
A packet is transmitted when the number of tokens in the token bucket is sufficient to allow transmission. After a packet is transmitted, the number of tokens decreases accordingly. The details are as follows:

1. Incoming packets are categorized into different traffic classes. If the packets belong to a traffic class for which the rate is limited, for example, packets 2, 3, and 4 in **Figure 15-23**, the packets are sent to the token bucket for processing. If the packets do not belong to any traffic class, for example, packet 1 in **Figure 15-23**, the rate of the packets is not limited and the packets are transmitted directly.
2. If the token bucket stores sufficient tokens, packets are transmitted, for example, packets 2 and 3 in **Figure 15-23**.
3. If the token bucket stores insufficient tokens, packets are dropped, for example, packet 4 in **Figure 15-23**.
4. The system places tokens in the token bucket at a user-defined rate. When new tokens are generated in the bucket, successive packets can be transmitted.

When the token bucket is full, the system transmits all packets that are the same size as the tokens in the bucket. This process allows for burst transmission. When the token bucket is empty, the system cannot transmit any packet. The system resumes transmissions only after new tokens have been generated. This means that the traffic transmission rate is consistently lower than or equal to the token generation rate, so as to achieve the rate limiting goal.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-23 CAR implementation using a single token bucket (process in clockwise direction)



In the figure above, the digits 1, 2, 3, and 4 are the numbers of packets. After a packet passes through the bucket, the number of tokens decreases by the size of the packet. For easy understanding, we suppose that all packets have the same size.

Principle of Dual Token Buckets (trTCM)

The two rate three color marker (trTCM) algorithm defined by RFC2698 is used for traffic policing and marking to achieve effective bandwidth management. If static bandwidth is planned for a network, the trTCM algorithm can at least ensure the basic bandwidth (CIR) for users, and allows users to obtain extra bandwidth (PIR) when the network has sufficient bandwidth. In this way, network resources are used more efficiently.

trTCM principles are as follows:

- trTCM uses two token buckets (P and C buckets). The maximum capacity of the P bucket is represented by the peak burst size (PBS) and that of the C bucket is represented by the committed burst size (CBS). The two token buckets are initially full.

  NOTE

  PBS must be larger than CBS.

- After a packet passes through the tokens, the number of tokens in the buckets decreases accordingly.

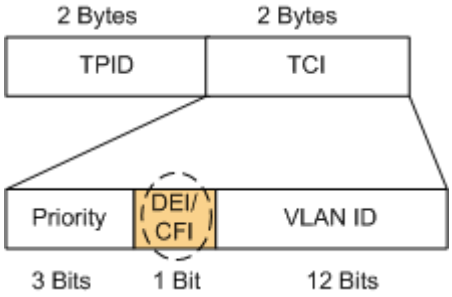
 Copyright © Huawei Technologies Co., Ltd.

- The system generates tokens in the P bucket at the PIR and in the C bucket at the CIR (PIR >= CIR) per second. The total tokens in the buckets are always less than maximum capacity.
- Based on the size of tokens in the buckets, the system marks the DEI bit (as shown in **Figure 15-24**) in an incoming packet with a corresponding color (green, yellow, or red). Color coded packets help the system to prevent or manage congestion during data processing.

  NOTE

  The MA5800 uses the CFI bit defined by 802.1q as the DEI.

Figure 15-24 DEI bit defined by 802.1ad (same as the CFI bit defined by 802.1q)

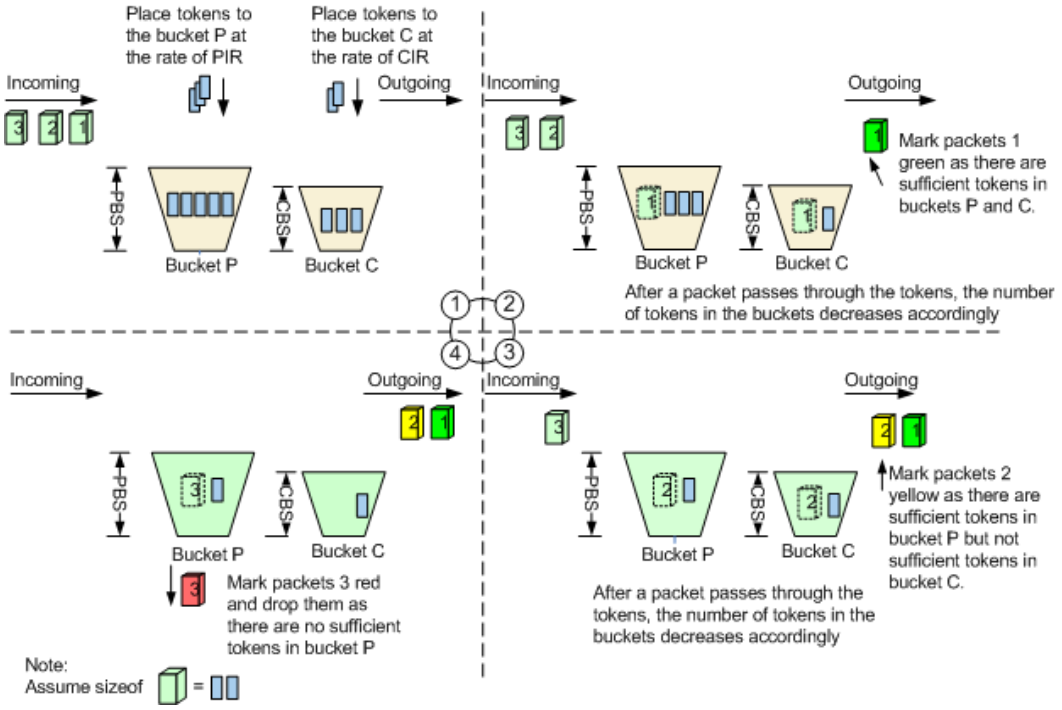


As shown in **Figure 15-25**:

1. If a packet does not exceed CIR, it is marked green (the DEI bit in the packet is set to 0) and is allowed to pass. Packet 1 in the figure is an example.
2. If a packet exceeds CIR but does not exceed PIR, it is marked yellow (the DEI bit in the packet is set to 1) and is allowed to pass. Packet 2 in the figure is an example.
3. If a packet exceeds PIR, the packet is marked red and is directly dropped. Packet 3 in the figure is an example.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-25 CAR implementation using the trTCM algorithm (process in clockwise direction)



In the figure above, the digits 1, 2, and 3 are the numbers of packets. After a packet passes through the bucket, the number of tokens decreases by the size of the packet. For easy understanding, we suppose that all packets have the same size.

Enhanced trTCM: Coloring Packets Based on CAR Threshold

The trTCM algorithm ensures that different types of packets are all forwarded when they do not exceed CIR and are forwarded fairly when they reach PIR. The trTCM algorithm policies packets based on the bandwidth and it does not mark colors for packets based on priorities. Therefore, the trTCM algorithm cannot guarantee bandwidth for high-priority services. The enhanced trTCM algorithm can resolve the issue of the trTCM algorithm because it marks colors for packets based on CAR thresholds.

The implementation principle is as follows:

- When the enhanced trTCM algorithm is implemented for equipment, users can set different CAR thresholds for different packets by running the **car-threshold** command, ensuring that a high CAR threshold is set for high-priority packets.

  NOTE

  The CAR threshold is the percentage by which tokens in the C and P buckets have decreased. The percentage of the remaining tokens in the buckets is obtained by subtracting the CAR threshold from 1.

- When packets of different priorities sequentially pass the P bucket and C bucket:

 Copyright © Huawei Technologies Co., Ltd.

  - If a packet of a certain priority in the P bucket exceeds the CAR threshold, the packet is directly dropped. Otherwise, the packet enters the C bucket.
  - If a packet of a certain priority in the C bucket exceeds the CAR threshold, the packet is marked yellow; otherwise, it is marked green.
- After a packet passes through the tokens, the number of tokens in the buckets decreases accordingly.

As shown above, enhanced trTCM ensures that packets with higher priorities are marked green and their bandwidths are guaranteed.

As shown in **Figure 15-26**, there are two token buckets C and P with depths CBS and PBS respectively. The two buckets have token counts Tc and Tp, respectively. Tp(t) and Tc(t) represent the number of tokens in the P and C buckets respectively at time t. Initially (t = 0), the P and C buckets are full, that is, Tp(0) = PBS and Tc(0) = CBS. Then, Tp increases by one for PIR times per second until reaching PBS and Tc increases by one for CIR times per second until reaching CBS. Lvl(i) represents the percentage of the remaining tokens corresponding to a specific priority.

When an i-priority packet with length of B bytes arrives at time t:

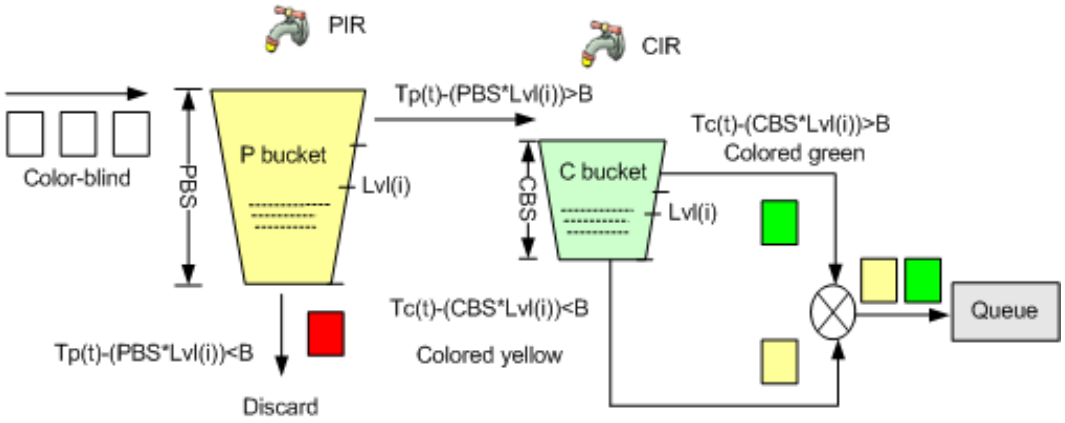
1. If Tp(t) - (PBS*Lvl(i)) < B, the device drops the packet; otherwise, the device sends the packet to the C bucket, and Tp(t) decreases by B bytes.
2. If Tc(t) - (CBS*Lvl(i)) < B, the device marks the packet yellow and Tc(t) decreases by B bytes; otherwise, the device marks the packet green and Tc(t) decreases by B bytes.

For example, when the CAR threshold is 12% for 0-priority packets and is 100% for 6-priority packets, the enhanced trTCM algorithm enables the device to behave as follows:

- When the incoming packets enter the P bucket at a certain time, if the percentage of the remaining tokens in the bucket is less than 88%, the device drops 0-priority packets and sends 6-priority packets to the C bucket. When the incoming packets enter the C bucket at a certain time, if the percentage of the remaining tokens in the bucket is less than 88%, the device marks 0-priority packets yellow and 6-priority packets green.
- When the incoming packets enter the P bucket at a certain time, if there is no token in the bucket, the device also drops 6-priority packets. When the incoming packets enter the C bucket at a certain time, if there is no token in the bucket, the device drops 0-priority packets and marks 6-priority packets yellow.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-26 Coloring packets based on CAR threshold



15.7.4 Traffic Policing Mode

As shown in **Table 15-5**, the MA5800 supports multiple traffic policies:

- Rate limiting based on service flows
- Rate limiting based on port+priority
- Rate limiting based on GEM port+CoS group
- Rate limiting based on service flow+priority
- Rate limiting based on groups of service flows
- Rate limiting based on port+VLAN
- Rate limiting based on priority group
- Rate limiting based on priority groups of service flow groups
- Rate limiting based on ONT
- Rate limiting based on Ethernet port on ONT

 Copyright © Huawei Technologies Co., Ltd.

Table 15-5 Traffic policing mode and token bucket algorithm

Traffic Policy	Description	Token Bucket Algorithm
Rate limiting based on service flows	When configuring service flows, you can limit the rate of the service flows by binding upstream and downstream service flows to an upstream and downstream profile, respectively. If there is no limit on the rate, you can configure the upstream and downstream profiles as **car-off**. Rate limiting based on service flows and rate limiting based on port+CoS cannot be applied to the same board or port at the same time. The configuration commands are: • **car-mode service-port** • **service-port** • and **traffic table ip**.	trTCM
Rate limiting based on port +priority	Rate limiting based on GEM port+CoS is used for GPON applications. Rate limiting based on service flows and rate limiting based on port+CoS cannot be applied to the same board or port at the same time. The configuration commands are: • **car-mode port-cos** • **car-port *portid* cos** • **car-port** *portid* **ont** *ontid* **gemindex** *gemindex* and **cos**.	trTCM

 Copyright © Huawei Technologies Co., Ltd.

Traffic Policy	Description	Token Bucket Algorithm
Rate limiting based on GEM port+CoS group	One GEM port can carry services with different priorities. After the priority of the specific service is added to the CoS group, rate limiting is performed based on this CoS group. This meets the carrier requirement that rates of various services carried on the same GEM port are limited. The configuration commands are: • **car-mode port-cos** • **cos-group-table** • **car-port** *portid* **ont** *ontid* **gemindex** *gemindex* **cos-group-table** *table-index* • **car-port** *portid* **ont** *ontid* **gemindex** *gemindex* **cos-group** and *group-id*.	trTCM
Rate limiting based on service flow +priority	When configuring service flows, you can limit the rate of the service flows by binding upstream and downstream service flows to an upstream and downstream profile, respectively. If there is no limit on the rate, you can configure the upstream and downstream profiles as **car-off** (system default). The configuration command is: **car-threshold profile**.	Enhanced trTCM
Rate limiting based on groups of service flows	You can bind multiple upstream or downstream service flows into one group and apply the same CAR to the group. The configuration command is: **car-group**.	trTCM
Rate limiting based on port +VLAN	You can limit the rates for packets with specified VLANs on specified ports. Only the Ethernet board supports this traffic policy. The configuration commands are: **car-port** *portid* and **vlan**.	trTCM

 Copyright © Huawei Technologies Co., Ltd.

Traffic Policy	Description	Token Bucket Algorithm
Rate limiting based on priority group	In this rate limiting, priority groups are of the same service flow. Priorities are grouped based on inner and outer 802.1p tags. The configuration commands are: • **priority-group profile** • and **car-group**.	trTCM

Traffic Policy	Description	Token Bucket Algorithm
Rate limiting based on priority groups of service flow groups	Create service flow groups first. Then, run the **car-group priority group profile** command to enable priority group CAR of service flow groups. Run the **car-group** command to configure the **traffic-table** parameter and apply the priority-group profile to achieve rate limiting based on priority groups of service flow groups. The configuration commands are: ● **car-group add-member service-port** ● **car-group priority group profile** ● and **car-group**.	TrTCM Description: ● Priority group CAR of service flow groups, and CAR of service flows or priority group CAR of service flows are mutually exclusive. That is, after priority group CAR of service flow groups is configured, CAR of service flows or priority group CAR of service flows cannot be configured. On the contrary, after CAR of service flows or priority group CAR of service flows is configured, priority group CAR of service flow groups cannot be configured. CAR of service flows and priority group CAR of service flows are not mutually exclusive. ● Connection-oriented service flows can be added to groups based on source and destination. ● CAR of service flows and priority group CAR of service flow groups support specification sharing. The shared CAR specification is twice the CAR

 Copyright © Huawei Technologies Co., Ltd.

Traffic Policy	Description	Token Bucket Algorithm
		specification of service flows on boards. For example, if the CAR specification of service flows on boards is 16K, the shared CAR specification is 32K.
Rate limiting based on ONT	You can limit the rates for downstream packets of a specified ONU. The configuration command is: **traffic-limit ont**.	Single token bucket Note: This rate limiting policy applies when the single token bucket algorithm and the rate limiting policy is bound to the trTCM traffic profile. In the trTCM traffic profile, CIR must be equal to PIR; otherwise, the preconfigured PIR will supersede the specified rate limiting policy.
Rate limiting based on Ethernet port on ONT	You can limit the rates for upstream and downstream packets of specified Ethernet ports on an ONT. The configuration command is: **ont port car**.	Single token bucket Note: This rate limiting policy applies when the single token bucket algorithm and the rate limiting policy is bound to the trTCM traffic profile. In the trTCM traffic profile, CIR must be equal to PIR; otherwise, the preconfigured CIR will supersede the specified rate limiting policy.

 Copyright © Huawei Technologies Co., Ltd.

Traffic Policy	Description	Token Bucket Algorithm
Traffic suppression	Using this policy, the system suppresses broadcast, unknown multicast, and unknown unicast packets for inbound ports. The traffic suppression prevents these packets from consuming excessive network resources and therefore protects the network from congestion. The configuration command is: **traffic-suppress**.	-

15.7.5 Configuring the Traffic Policing

15.7.5.1 Configuring Rate Limitation Based on Service Port

This topic describes how to limit rate on a specific service flow (on behalf of a different type of service or users) through the traffic parameters defined in the IP traffic profile.

Context

- The system has seven default IP traffic profiles with the IDs of 0-6. You can run the **display traffic table** command to query the traffic parameters of the default traffic profiles.
- It is recommended that you use the default traffic profiles. A new IP traffic profile is created only when the default traffic profiles cannot meet the requirements.

Procedure

Step 1 Configure the rate limitation mode to be service-port.

In the service board mode, run the **car-mode** [*portlist*] **service-port** command to set the rate limitation mode of board or port to be service-port.

Step 2 Configure the parameters of IP traffic profile.

When the default traffic profiles cannot meet the requirements, run the **traffic table ip** command to configure the IP traffic profile.

Table 15-6 shows the traffic parameters related to rate limitation in IP traffic profile.

 Copyright © Huawei Technologies Co., Ltd.

Table 15-6 Traffic parameters and parameter Description

Item	Parameter Description
cir *cir*	Stands for committed information rate. The unit is bit/s. CIR is mandatory, and must be an integer multiple of 64. If it is not, the value is rounded down to a nearest integer multiple of 64 but cannot be smaller than 64.
cbs *cbs*	Stands for committed burst size. It indicates the maximum capacity that a token bucket buffers tokens. The unit is byte. CBS is optional. If the parameter is not specified, it can be obtained by the formula min(2000 + 32*cir,10240000).
pir *pir*	Stands for peak information rate. It provides users the maximum bandwidth when the system is idle. The unit is bit/s. PIR is optional. If the parameter is not specified, it can be obtained by the formula min(2*cir, 10240000). The pir cannot be smaller than cir.
pbs *pbs*	Stands for peak burst size. It indicates the maximum capacity that a token bucket buffers tokens. The unit is byte. PBS is optional. If the parameter is not specified, it can be obtained by the formula min(2000+32*pir, 10240000).

NOTE

The system marks the service packets with colors according to the CIR and PIR parameters. To be specific, for the packets whose rate is equal to or lower than CIR, the system marks them as green (allowed to pass). For the packets whose rate is higher than CIR and lower than PIR, the system marks them as yellow (allowed to pass). For the packets whose rate is higher than PIR, the system marks them as red, and drops such packets.

Step 3 Run the **service port** command to bind the specified IP traffic profile.

----End

Example

Assume that the CIR is 2048 Kbit/s, 802.1p priority of the outbound packet is 6, and the scheduling policy of the inbound packet is Tag-In-Package. To add traffic profile 9 with these settings, do as follows:

```
huawei(config)#traffic table ip index 9 cir 2048 priority 6 priority-policy tag-In-Package
  Create traffic descriptor record successfully
  ------------------------------------------------
  Traffic Table Index          : 9
  Traffic Table Name           : ip-traffic-table_9
  Specified Outer-Priority     : 6
  Outer-Priority Mapping Source: -
  Outer-Priority Mapping Index : -
  Default Outer-Priority       : -
  Specified Inner-Priority     : -
  Inner-Priority Mapping Source: -
  Inner-Priority Mapping Index : -
  Default Inner-Priority       : -
  Specified DSCP               : -
```

 Copyright © Huawei Technologies Co., Ltd.

```
  DSCP Mapping Source          : -
  DSCP Mapping Index           : -
  CIR                          : 2048 kbps
  CBS                          : 67536 bytes
  PIR                          : 4096 kbps
  PBS                          : 133072 bytes
  Fix                          : 0 kbps
  CAR Threshold Profile        : -
  Color Mode                   : color-blind
  Coupling Flag                : enable
  Rate Profile Index           : -
  Rate Profile Name            : -
  Color Policy                 : dei
  Color DEI                    : unmark-dei
  Color Source                 : outer-dei
  Drop Precedence              : remarked-outer-priority
  Priority Policy              : tag-pri
  Que-Policy Profile ID        : -
  Pri-Group Profile ID         : -
  Pri-Group Profile Name       : -
  Referenced Status            : not used
  E2E Mapping Priority         : -
  E2E Mapping Index            : -
  E2E Default Priority         : -
  Yellow WRR Weight            : auto
  Green WRR Weight             : auto
  ------------------------------------------------
huawei(config)#display traffic table ip index 9
  ------------------------------------------------
  Traffic Table Index          : 9
  Traffic Table Name           : ip-traffic-table_9
  Specified Outer-Priority     : 6
  Outer-Priority Mapping Source: -
  Outer-Priority Mapping Index : -
  Default Outer-Priority       : -
  Specified Inner-Priority     : -
  Inner-Priority Mapping Source: -
  Inner-Priority Mapping Index : -
  Default Inner-Priority       : -
  Specified DSCP               : -
  DSCP Mapping Source          : -
  DSCP Mapping Index           : -
  CIR                          : 2048 kbps
  CBS                          : 67536 bytes
  PIR                          : 4096 kbps
  PBS                          : 133072 bytes
  Fix                          : 0 kbps
  CAR Threshold Profile        : -
  Color Mode                   : color-blind
  Coupling Flag                : enable
  Rate Profile Index           : -
  Rate Profile Name            : -
  Color Policy                 : dei
  Color DEI                    : unmark-dei
  Color Source                 : outer-dei
  Drop Precedence              : remarked-outer-priority
  Priority Policy              : tag-pri
  Que-Policy Profile ID        : -
  Pri-Group Profile ID         : -
  Pri-Group Profile Name       : -
  Referenced Status            : not used
  E2E Mapping Priority         : -
  E2E Mapping Index            : -
  E2E Default Priority         : -
  Yellow WRR Weight            : auto
  Green WRR Weight             : auto
  ------------------------------------------------
```

 Copyright © Huawei Technologies Co., Ltd.

15.7.5.2 Configuring Rate Limitation Based on Port+CoS

This topic describes how to configure Rate limiting based on port+CoS so that different IP traffic profiles can be specified for the traffic streams that have different 802.1p priorities on a port.

Prerequisites

The IP traffic profile is configured. For details, see **15.7.5.1 Configuring Rate Limitation Based on Service Port**.

Context

- Rate limiting based on service ports conflicts with Rate limiting based on port +CoS. By default, the system supports Rate limiting based on service ports.
- If service ports are configured on the board, the rate limiting mode of the board cannot be changed.

Procedure

Step 1 Configure the rate limitation mode to be port-cos.

In the service board mode, run the **car-mode** [*portlist*] **port-cos** command to configure the rate limiting mode of the service board to rate limiting based on port+CoS.

Step 2 Configure rate limitation for service flows with specified 802.1p priority.

Run the **car-port** command to specify the 802.1p priority for the port, and bind an IP traffic profile to the traffic streams that meet the specified 802.1p priority.

When rate limiting based on port+CoS is selected for a board, pay attention to the following points:

- For a non-xPON board, you can bind the corresponding traffic profile in the inbound/outbound direction according to a CoS value of a port on the board.
- For a GPON board, you can bind the corresponding traffic profile in the inbound/outbound direction according to a CoS value of a GEM port on the board.

----End

Example

To configure GEM port 1 on ONT 0 of the board in port 0/2/0, and bind traffic profile 2 to the packets with priority 5, do as follows:

```
huawei(config)#interface gpon 0/2
huawei(config-if-gpon-0/2)#car-mode port-cos
huawei(config-if-gpon-0/2)#car-port 0 ont 0 gemindex 1 cos 5 inbound 2 outbound 2
huawei(config-if-gpon-0/2)#display car-mode
  The CAR mode of the board: port-cos
 ------------------------------------------------
```

15.7.5.3 Configuring User-based Rate Limitation

In the user-based rate limitation, the VoIP, IPTV service, and Internet access service of each user share a total user bandwidth. When there is no voice or IPTV service,

 Copyright © Huawei Technologies Co., Ltd.

the Internet access service can hold a burst of the total user bandwidth so that the total user bandwidth can be managed in a unified manner.

Context

When the user uses the Triple play service, the VoIP, IPTV service, and Internet access service of each user share a total user bandwidth. All services of the user hold the total user bandwidth, and the service with the highest CoS priority is ensured first. When other services carry no traffic, each service can hold a burst of the total user bandwidth. The multicast bandwidth is determined by the bandwidth of demanded programs. The total bandwidth of demanded programs cannot exceed the total user bandwidth. The bandwidth of voice services is fixed.

NOTE

PON multicast services do no support rate limitation in CAR-group mode.

Procedure

- For PON access users.
  - In the user-based rate limitation, multiple service ports of a user are added to a rate-limited group. Through the QoS strategy applied on the rate-limited group, the total user bandwidth is ensured on the basis that the committed information rate (CIR) and peak information rate (PIR) of each service are ensured, and each service is allowed to hold a burst of the total user bandwidth.
  - Only the GPON service boards support user-based rate limitation.

  a. Run the **traffic table ip** command to create an IP traffic profile to configure the CoS priority of each service and ensure the CIR and PIR.

     - The CoS priorities of services are VoIP, IPTV service, and Internet access service in a descending order.
     - In the IP traffic profile used by the rate-limited group, the PIR must be equal to or larger than the sum of CIRs of all services in other IP traffic profiles.

  b. Run the **service-port** command to create service ports of the VoIP, IPTV service, and Internet access service, using the IP traffic profile created in step 1.

  c. Run the **car-group** command to create the rate-limited group of service ports to manage the total user bandwidth of multiple services.

     - To ensure the user bandwidth, the PIR of the rate-limited group must be equal to or larger than the sum of CIRs of all services in the rate-limited group.
     - The PIR is equal to the total user bandwidth. In the case that any two services carry no traffic, the third service can hold a burst of the total user bandwidth.

  d. Run the **car-group add-member service-port** command to add service ports to the rate-limited group.

     Pay attention to the following points when adding service ports to the rate-limited group:

 Copyright © Huawei Technologies Co., Ltd.

- Only service ports of the same PON port can be added to the same rate-limited group.
- For Type C and Type D, only service ports of the same ONT can be added to the same rate-limited group.
- One service port cannot be added to multiple rate-limited groups.
- A maximum of eight service ports can be added to a rate-limited group.

----End

Example

Assume that under GPON port 0/2/1, the user with the ONT 1 is provided with the VoIP, IPTV, and Internet access services. Set the total user bandwidth to 10 Mbit/s, add rate-limited group 0, add service ports 100, 101, and 102 of the user to rate-limited group 0, and use traffic profile 30 to control traffic of rate-limited group 0. In the case that any two services carry no traffic, the third service can hold a burst of the total user bandwidth. To perform such a configuration with the following parameters, do as follows:

- Service port 100 of the Internet access service uses traffic profile 10, with the CIR 2 Mbit/s and the 802.1p priority 4.
- Service port 101 of the VoIP service uses traffic profile 11, with the CIR 1 Mbit/s and the 802.1p priority 6.
- Service port 102 of the IPTV service uses traffic profile 12, with the packet rate not limited and the 802.1p priority 5.

```
huawei(config)#traffic table ip index 10 cir 2048 pir 10240 priority 4 priority-policy local-Setting
huawei(config)#service-port 100 vlan 2 gpon 0/2/1 ont 1 gemport 4 multi-service user-vlan 20 rx-cttr 10 tx-cttr 10
huawei(config)#traffic table ip index 11 cir 1024 pir 10240 priority 6 priority-policy local-Setting
huawei(config)#service-port 101 vlan 2 gpon 0/2/1 ont 1 gemport 5 multi-service user-vlan 30 rx-cttr 11 tx-cttr 11
huawei(config)#traffic table ip index 12 cir off priority 5 priority-policy local-Setting
huawei(config)#service-port 102 vlan 2 gpon 0/2/1 ont 1 gemport 6 multi-service user-vlan 40 rx-cttr 12 tx-cttr 12
huawei(config)#traffic table ip index 30 cir 10240 pir 10240 priority 3 priority-policy local-Setting
huawei(config)#car-group 0 inbound traffic-table index 30 outbound traffic-table index 10
huawei(config)#car-group 0 add-member service-port 100-102
huawei(config)#display car-group 0

Command:
        display car-group 0
  ------------------------------------------------------------------------------
  GroupID                                 Member List     RX   TX
  ------------------------------------------------------------------------------
      0                                                   10   30
  ------------------------------------------------------------------------------
  Note: S--Source of connection-oriented flow,
        D--Destination of connection-oriented flow
```

15.7.5.4 Configuring Rate Limitation Based on Port+VLAN

After configuring rate limitation based on port+VLAN, you can specify different IP traffic profiles for different VLAN packets carried on the same port.

 Copyright © Huawei Technologies Co., Ltd.

Prerequisites

- A proper IP traffic profile must be created and the index of the IP traffic profile to be used must be confirmed. For details about the configuration method, see **15.7.5.1 Configuring Rate Limitation Based on Service Port**.
- Currently, only the ETH board supports rate limitation based on port+VLAN.

Procedure

Step 1 In the global config mode, run the **interface eth** command to enter the ETH mode.

Step 2 Run the **car-port portid vlan** command to configure rate limitation based on port +VLAN.

This command can be used to configure IP traffic profiles for the packets in the specified VLAN range on the specified port, implementing inbound and outbound rate limitation.

----End

Example

To configure port 0 on the ETH board in slot 0/9/0, and use traffic profile 6 for controlling the packets with VLAN 10, do as follows:

```
huawei(config)#display traffic table ip index 6
  ---------------------------------------------------
  Traffic Table Index          : 6
  Traffic Table Name           : ip-traffic-table_6
  Specified Outer-Priority     : 0
  Outer-Priority Mapping Source: -
  Outer-Priority Mapping Index : -
  Default Outer-Priority       : -
  Specified Inner-Priority     : 0
  Inner-Priority Mapping Source: -
  Inner-Priority Mapping Index : -
  Default Inner-Priority       : -
  Specified DSCP               : -
  DSCP Mapping Source          : -
  DSCP Mapping Index           : -
  CIR                        : off
  CBS                        : off
  PIR                        : off
  PBS                        : off
  Fix                        : off
  CAR Threshold Profile        : -
  Color Mode                 : color-blind
  Coupling Flag              : enable
  Rate Profile Index           : -
  Rate Profile Name            : -
  Color Policy               : dei
  Color DEI                  : unmark-dei
  Color Source               : outer-dei
  Drop Precedence              : remarked-outer-priority
  Priority Policy            : tag-pri
  Que-Policy Profile ID        : -
  Pri-Group Profile ID         : -
  Pri-Group Profile Name       : -
  Referenced Status          : not used
  E2E Mapping Priority         : -
  E2E Mapping Index            : -
  E2E Default Priority         : -
  Yellow WRR Weight            : auto
```

 Copyright © Huawei Technologies Co., Ltd.

```
  Green WRR Weight          : auto
  ------------------------------------------------
huawei(config)#interface eth 0/9
huawei(config-if-eth-0/9)#car-port 0 vlan 10  inbound 6 outbound 6
```

15.7.5.5 Configuring GPON Rate Limitation

This topic describes how to configure rate limitation for GPON services, thereby providing differentiated quality of service (QoS) for various GPON services.

Context

- There are multiple methods of rate-limiting GPON services, for example, rate-limiting downstream traffic by using an IP traffic profile and ACL rules, rate-limiting the ONT upstream bandwidth by using a DBA profile, and rate-limiting the GEM port and GEM port traffic on an ONT.
- Rate limitation on GPON services can be performed on the OLT and the ONT concurrently. If more than one rate limitation modes are configured in the system, the minimum rate prevails.
- Which method of rate-limiting the ONT upstream bandwidth is used depends on the ONT capability. Specifically, if an ONT supports various rate limitation methods and the ONT upstream traffic is small (for example, FTTH service), a DBA profile is a best choice to rate-limit the ONT upstream traffic. If a T-CONT carries upstream traffic for multiple users (for example, FTTB/FTTC service), rate limitation on GEM port is generally used to prevent a user from occupying bandwidth for a long time. If the priority of user packets is trustworthy (for example, an enterprise user), priority queue (PQ) scheduling is generally used.

Procedure

- Perform rate limitation on the OLT.
  - Rate limitation using an IP traffic profile includes two modes. For details, see **Configuring Traffic Management Based on Service Port**, and **Configuring Traffic Management Based on Port+CoS**.
  - Performing rate limitation by configuring an ACL rule can control the traffic matching the ACL rule. For details, see **Controlling the Traffic Matching an ACL Rule**.
- Perform rate limitation on the ONT.
  a. Run the **dba-profile add** command to add a DBA profile. The DBA profile is used to schedule the ONT upstream bandwidth properly, achieving the best bandwidth utilization.

     A DBA profile supports five types (Type1 to Type5). Generally, services with higher priorities adopt Type1 or Type2 DBA profiles and services with lower priorities adopt Type3 or Type4 DBA profiles. **Table 15-7** shows the features of the DBA profile of each type.

 Copyright © Huawei Technologies Co., Ltd.

Table 15-7 The features of the DBA profile

Profile Type	Features
Type1	Indicates the fixed bandwidth. After the DBA profile of Type1 is bound, the system assigns a specified bandwidth, regardless of whether there is upstream traffic.
Type2	Indicates the assured bandwidth. After the DBA profile of Type2 is bound, the system meets the bandwidth requirements if the upstream traffic does not exceed a specified value. When there is no upstream traffic, the system does not assign any bandwidth.
Type3	Indicates the hybrid of assured bandwidth and non-assured bandwidth. The DBA profile of Type3 specifies an assured value and non-assured value. After assigning the fixed bandwidth and assured bandwidth, the system assigns the remaining bandwidth (if any) to the user bound with the DBA profile of Type3 (the assigned bandwidth does not exceed the non-assured bandwidth).
Type4	Indicates the best-effort bandwidth. The DBA profile of Type4 just specifies a maximum value. After the DBA profile of Type4 is bound, its priority for obtaining the bandwidth is the lowest. That is, after assigning the fixed bandwidth, assured bandwidth, and non-assured bandwidth, the system assigns the remaining bandwidth (if any) to the user bound with the DBA profile of Type4 (the assigned bandwidth does not exceed the maximum value).
Type5	Indicates the hybrid bandwidth. The preceding four types of values need to be specified.

b. Run the **ont-lineprofile gpon** command to add a GPON ONT line profile, and then enter the GPON ONT line profile mode.

c. Run the **tcont** command to bind a T-CONT to the DBA profile.

   It is recommended that one service type use one T-CONT and different T-CONTs be planned with different bandwidth assurance types.

d. Run the **qos-mode** command to configure a QoS mode of the GPON ONT line profile to ensure that the QoS mode is the same as that of the GEM port.

   By default, the QoS mode of the GPON ONT line profile (that is, the ONT scheduling mode) is priority queue (PQ). The QoS mode includes:

   - gem-car: Indicates the rate limitation mode based on the GEM port of the T-CONT. Rate limitation is performed on a specified GEM port in the ONT upstream direction. To select the gem-car mode, set **gem add** to **gem-car**. The maximum traffic is determined by the DBA profile bound to the GEM port. If a T-CONT contains multiple GEM

 Copyright © Huawei Technologies Co., Ltd.

ports, the scheduling mechanism of packets between multiple GEM ports depends on the default scheduling mechanism of the ONT.

- flow-car: Indicates the rate limitation mode based on traffic streams of a GEM port. Rate limitation is performed on a specified traffic stream in the ONT upstream direction. To select the flow-car mode, set **gem mapping** to **flow-car**. The maximum traffic is determined by the DBA profile bound to the traffic stream. Flow-car is more specific than gem-car. After rate limitation based on traffic streams is performed, traffic is scheduled in the T-CONT queue. The scheduling mechanism depends on the default scheduling mechanism of the ONT. Before configuring flow-car, make sure that the required traffic profile is created by running the **traffic table ip** command.

  NOTE

  The traffic stream in this topic refers to the service channel between an ONT and OLT. It is different the service port created by running the **service-port** command.

- priority-queue: Indicates the PQ mode based on the GEM port of the T-CONT. Traffic is scheduled based on PQ between multiple GEM ports in the ONT upstream direction. To select priority-queue mode, set **gem add** to **priority-queue**. By default, the system supports eight (0–7) queues. Queue 7 has the highest priority and services of queue 7 are preferentially scheduled. The maximum traffic is determined by the DBA profile to which the T-CONT is bound.

e. Run the **commit** command to make the profile configuration take effect. The configuration of the line profile takes effect only after you run this command.

----End

Example

Assume that:

- A user under ONT 1 connected to GPON port 0/2/1 requires 2 Mbit/s high-speed Internet access service.
- The priority of user packets is trustworthy. The PQ scheduling mechanism is used, with priority 1.
- The default IP traffic profile, namely IP traffic profile 5 is used for rate limitation on a GPON port, with CIR of 2048 Kbit/s.
- DBA profile 10 of Type4 is used and the maximum bandwidth in the ONT upstream direction is 100 Mbit/s.

To perform the preceding configurations, do as follows:

```
huawei(config)#dba-profile add profile-id 10 type4 max 102400
huawei(config)#ont-lineprofile gpon profile-id 5
huawei(config-gpon-lineprofile-5)#tcont 1 dba-profile-id 10
huawei(config-gpon-lineprofile-5)#qos-mode Priority-queue
huawei(config-gpon-lineprofile-5)#gem add 1 eth tcont 1 priority-queue 1
huawei(config-gpon-lineprofile-5)#mapping-mode vlan
huawei(config-gpon-lineprofile-5)#gem mapping 1 2 vlan 10
huawei(config-gpon-lineprofile-5)#commit
huawei(config-gpon-lineprofile-5)#quit
huawei(config)#interface gpon 0/2
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-if-gpon-0/2)#ont confirm 1 ontid 1 sn-auth 32303131B39FD641  snmp ont-lineprofile-id
5
huawei(config-if-gpon-0/2)#quit
huawei(config)#service-port 101 vlan 100 gpon 0/2/1 ont 1 gemport 1 rx-cttr 5 tx-cttr 5
```

15.7.5.6 Configuring Rate Limiting Based on Priority Group

By configuring rate limiting based on priority group, you can specify different IP traffic profiles for 802.1p groups over a service flow.

Procedure

Step 1 In global config mode, run the **rate profile** command to create a rate profile.

Step 2 In rate profile mode, run the **cir** command to configure rate limiting parameters and then exit rate profile mode.

Step 3 In global config mode, run the **priority-group profile** command to enter priority group profile mode and create a priority group profile.

Step 4 In priority group profile mode, run the **priority-source(priority-group profile)** and **car-group(priority-group profile)** commands to configure the source of the priority group and specify the priority group rate profile.

Step 5 Run the **traffic table ip** command to bind the priority group profile.

----End

Example

To create 2 priority groups by using the outer priority of packets and rate profile 1 for rate limiting, do as follows:

```
huawei(config)#rate profile
huawei(config-rate-profile-1)#cir 64
huawei(config-rate-profile-1)#quit
huawei(config)#priority-group profile
huawei(config-priority-group-profile-1)#priority-source outer-8021p cos-group-table 2 0
huawei(config-priority-group-profile-1)#car-group 0 rate profile profile-id 1
huawei(config-priority-group-profile-1)#car-group 1 rate profile profile-id 1
huawei(config-priority-group-profile-1)#quit
huawei(config)#traffic table ip index 1000 rate profile profile-id 1 priority user-co
s mapping-profile 0 0 priority-policy local-Setting priority-group profile profi
le-id 1
  Create traffic descriptor record successfully
  ------------------------------------------------------------
  Traffic Table Index          : 1000
  Traffic Table Name           : ip-traffic-table_1000
  Specified Outer-Priority     : -
  Outer-Priority Mapping Source: user-cos
  Outer-Priority Mapping Index : 0
  Default Outer-Priority       : 0
  Specified Inner-Priority     : -
  Inner-Priority Mapping Source: -
  Inner-Priority Mapping Index : -
  Default Inner-Priority       : -
  Specified DSCP               : -
  DSCP Mapping Source          : -
  DSCP Mapping Index           : -
  CIR                          : -
  CBS                          : -
  PIR                          : -
  PBS                          : -
```

 Copyright © Huawei Technologies Co., Ltd.

```
Fix                        : -
CAR Threshold Profile      : -
Color Mode                 : -
Coupling Flag              : -
Rate Profile Index         : 1
Rate Profile Name          : rate-profile_1
Color Policy               : dei
Color DEI                  : unmark-dei
Color Source               : outer-dei
Drop Precedence            : remarked-outer-priority
Priority Policy            : local-pri
Que-Policy Profile ID      : -
Pri-Group Profile ID       : 1
Pri-Group Profile Name     : priority-group-profile_1
Referenced Status          : not used
E2E Mapping Priority       : -
E2E Mapping Index          : -
E2E Default Priority       : -
Yellow WRR Weight          : auto
Green WRR Weight           : auto
-------------------------------------------------------------
```

15.8 Traffic Shaping

15.8.1 Introduction to Traffic Shaping

Definition

Traffic shaping is to shape irregular traffic or traffic that does not comply with preset traffic features to realize bandwidth matching between upstream and downstream networks.

Purpose

Control network traffic and prevent traffic bursts so that packets are transmitted at an even rate.

15.8.2 Basic Concepts of Traffic Shaping

Traffic shaping is generally implemented by using a buffer and a token bucket. When the network traffic is heavy, packets are first stored in the buffer, and then the buffered packets are sent at a specified rate through the token bucket.

Table 15-8 Basic concepts

Concept	Description
Buffer	A buffer is a space used to store packets. When tokens in the egress token bucket are used up, packets are stored in the buffer. Packets are discarded only when the buffer is full.

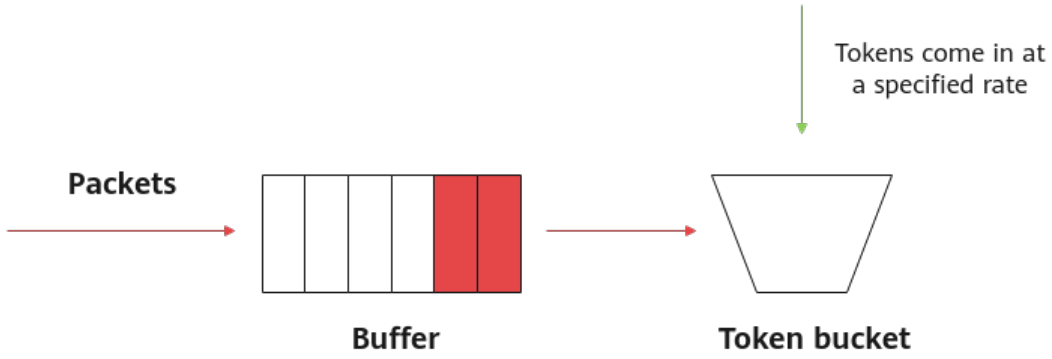
 Copyright © Huawei Technologies Co., Ltd.

Concept	Description
Token bucket	A token bucket is a container that stores tokens and is used to control data traffic. A token bucket allows burst data transmission while controlling the traffic. There are two types of token buckets: single bucket and dual buckets. Principle: A packet is transmitted only when there is a token of the equivalent size in the token bucket. After a packet is transmitted, the number of tokens of the equivalent size in the bucket decreases accordingly.

15.8.3 Traffic Shaping Implementation

This section describes the buffer function in traffic shaping. For token bucket principles, see **15.7.3 Implementation Principle: CAR**.

Figure 15-27 Traffic shaping implementation



Tokens are put into a token bucket at a specified rate. Packets can be transmitted only after obtaining tokens. When there is no token in the token bucket, the packets are not discarded. Instead, the packets are stored in the buffer and wait until there are tokens in the token bucket. Then, the buffered packets are sent. When the buffer is full, packets cannot be sent or stored in the buffer. In this case, packets are discarded.

Differences Between Traffic Shaping and Traffic Policing

Table 15-9 lists the differences between traffic shaping and traffic policing.

Table 15-9 Differences between traffic shaping and traffic policing

Traffic Policing	Traffic Shaping
Marks packets with priorities or discards packets in the case of burst traffic.	Stores packets in a buffer and discards packets only when the buffer is full in the case of burst traffic.

 Copyright © Huawei Technologies Co., Ltd.

Traffic Policing	Traffic Shaping
No extra buffer space is required.	Extra buffer space is required.
The packet loss rate is high.	The packet loss rate is reduced, but the latency is increased.
Jitter occurs during TCP rate limiting.	There is no jitter in TCP rate limiting.

15.8.4 Configuring Traffic Shaping

In traffic limiting scenarios, you can configure traffic shaping to reduce the packet loss rate and ensure that the traffic is even and stable.

15.8.4.1 Configuring a Rate Limit for an ONT

Traffic shaping can be used to limit the downstream traffic of an ONT.

Procedure

Step 1 Run the **traffic-limit ont** command to use an IP traffic profile to limit the traffic of downstream packets on an ONT. After this command is executed successfully, the traffic of downstream packets on the ONT is limited by traffic shaping. The packets that exceed the traffic limit are buffered to reduce packet loss.

NOTE

- You can select an IP traffic profile by name or SN.
- You can run the **display traffic table ip** command to query the IP traffic profiles in the system.
- You can run the **traffic table ip** command to create a new IP traffic profile.

Step 2 Run the **display traffic-limit ont** command to query the traffic limit information of the ONT. If the configured traffic limit is displayed, the configuration is successful.

----End

Example

To use IP traffic profile 6 to limit the traffic of downstream packets on ONT 1 connected to port 0/3/0, do as follows:

```
huawei(config)#traffic-limit ont
{ frameid/slotid/portid<S><Length 5-18> }:0/3/0
{ ontid<U><0,127> }:1
{ down-stream<K> }:down-stream
{ traffic-table<K> }:traffic-table
{ index<K>|name<K> }:index
{ index<U><0,1023> }:6

 Command:
        traffic-limit ont 0/3/0 1 down-stream traffic-table name onttraffictable

huawei(config)#display traffic-limit ont 0/3/0 1
```

 Copyright © Huawei Technologies Co., Ltd.

```
-------------------------------------------
F/ S/ P        ONT ID   DownTrafficProfile
-------------------------------------------
0/ 3/ 0        1        6
-------------------------------------------
```

15.8.4.2 Configure Queue-based Traffic Shaping for an ONT

Traffic shaping can be configured for ONT queue groups or queues to ensure even and stable egress traffic.

Procedure

Step 1 Run the **queue-shaping** command to configure traffic shaping and specify an IP traffic profile.

NOTE

- An IP traffic profile can be specified only by SN.
- You can run the **display traffic table ip** command to query the IP traffic profiles in the system.
- You can run the **traffic table ip** command to create a new IP traffic profile.

Step 2 Run the **display queue-shaping** command to check the traffic shaping configuration. If the configured information is displayed, the configuration is successful.

----End

Example

To configure traffic shaping for queue 1 on port 0/3/0 and specify IP traffic profile 15, do as follows:

NOTE

- The command formats for Ethernet access, xDSL access, and GPON access are slightly different. The following uses Ethernet access as an example.
- You can run this command to configure traffic shaping for a queue group and a queue. If **queue** is specified, traffic shaping is configured for a queue group. If **queue-list** is specified, traffic shaping is configured for a specified queue.

```
huawei(config)#queue-shaping
 { frameid/slotid/portid<S><Length 5-18>|frameid/slotid<S><Length 3-15> }:0/3/0
 { ont<K>|queue<K>|queue-list<K> }:queue-list
 { queue-list<S><Length 1-16> }:1
 { outbound<K> }:outbound { ip-traffic-table-index<U><0,1023> }:15

   Command:
        queue-shaping 0/3/0 queue-list 1 outbound 15

huawei#display queue-shaping 0/3/0
  ---------------------------------------------------------------------------
  F /S /P   QUEUE-ID1    QUEUE-ID2    OUTBOUND-INDEX       CIR        CBS
  ---------------------------------------------------------------------------
  0 /3 /0         3            4              15       6400      12800
  ---------------------------------------------------------------------------
```

 Copyright © Huawei Technologies Co., Ltd.

15.8.4.3 Configuring Rate Limit for an Ethernet Port

The traffic of a specified Ethernet port can be limited by traffic shaping. In traffic shaping mode, the packets exceeding the preset traffic limit are first stored in a buffer instead of directly discarded.

Procedure

Step 1 Run the **line-rate** command to configure the maximum rate in the inbound or outbound direction of a port.

NOTE

When **inbound** is specified for a port, the traffic shaping mode is used by default to limit the rate and buffer packets. If **inbound CAR** is specified, the port directly discards the packets that exceed the traffic limit.

Step 2 Run the **display qos-info** command to check the port configurations. If the configured information is displayed, the configuration is successful.

----End

Example

To limit the maximum rate of the inbound traffic transmitted from the user side to the network side on port 0/3/0 to 64 kbit/s, do as follows:

```
huawei(config)#line-rate
 { inbound<K>|integer<U><64,100000000>|outbound<K> }:inbound
 { car<K>|integer<U><64,100000000> }:64
 { port<K> }:port { frameid/slotid/portid<S><Length 5-18> }:0/3/0

    Command:
         line-rate inbound 64 port 0/3/0

huawei#display qos-info line-rate port 0/3/0
line-rate:
port 0/3/0:
   Inbound:
     line rate: 64Kbps
   Outbound:
     line rate: 128Kbps
```

15.9 Congestion avoidance

15.9.1 Introduction to Congestion avoidance

Definition

Congestion avoidance refers to a traffic control mechanism that monitors the utilization of network resources (such as queues or buffers), and drops packets to adjust the network traffic when congestion occurs, avoiding traffic overload on the network.

Congestion avoidance solves the issue of how to place packets in queue and how to drop packets.

 Copyright © Huawei Technologies Co., Ltd.

Purpose

It uses some algorithms to avoid worse congestion and utilizes network bandwidth.

15.9.2 Basic Concepts of Congestion avoidance

Congestion avoidance is implemented using packet drop algorithms. The MA5800 supports the following drop algorithms:

- Tail drop
- Priority-based early drop
- Color-based early drop
- Weighted random early discard (WRED)

Table 15-10 Basic concepts

Concept	Description	Effective Time
Tail drop	When a port queue is fully filled (to its maximum depth), the newly arriving packets are dropped until the queue has enough space to accept incoming traffic. Tail drop cannot ensure effective transmission of important data streams.	When the port queue is fully filled
Priority-based early drop	Packets with different priorities can be configured with different drop thresholds. Specifically, packets with higher priorities are configured with higher drop thresholds, while packets with lower priorities are configured with lower drop thresholds. When traffic congestion occurs on a port but the port queue is not fully filled, packets with higher priorities can enter queues that have a greater depth and are more burst-tolerant. In this way, these packets are less likely to be dropped. While packets with lower priorities are dropped preferentially because of the smaller queue depth.	When the port queue is partially filled
Color-based early drop	The system assigns different drop thresholds to different packets by marking the packets with different colors (yellow, green, or red; red packets are dropped directly) using the two rate three-color marker (trTCM) algorithm. In this way, when traffic congestion occurs on a port but the port queue is not fully filled, packets that do not exceed CIR (such packets are marked green) and packets that exceed CIR but do not exceed PIR (such packets are marked yellow) are forwarded; packets that exceed PIR (such packets are marked red) are dropped.	When the port queue is partially filled

 Copyright © Huawei Technologies Co., Ltd.

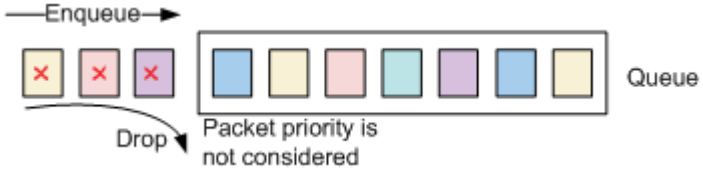
Concept	Description	Effective Time
WRED	Based on RED, WRED uses classification marks such as IP precedence, DSCP value, and MPLS EXP to differentiate drop policies. WRED effectively improves the quality of services with higher priorities.	When the port queue is partially filled

15.9.3 Implementation Principle

Tail Drop

When a port queue is fully filled (to its maximum depth), the newly arriving packets are dropped until the queue has enough space to accept incoming traffic, as shown in **Figure 15-28**.

Figure 15-28 Tail drop



Tail drop applies to all queues and cannot be manually configured.

Priority-based Early Drop

Packets with different priorities can be configured with different drop thresholds. Specifically, packets with higher priorities are configured with higher drop thresholds, while packets with lower priorities are configured with lower drop thresholds. When traffic congestion occurs on a port but the port queue is not fully filled, packets with higher priorities can enter queues that have a greater depth and are more burst-tolerant. In this way, these packets are less likely to be dropped. While packets with lower priorities are dropped preferentially because of the smaller queue depth.

Priority-based early drop applies to scenarios in which packets with different priorities enter the same queue.

The **early-drop** command can be used to configure the early drop thresholds for packets of different priorities.

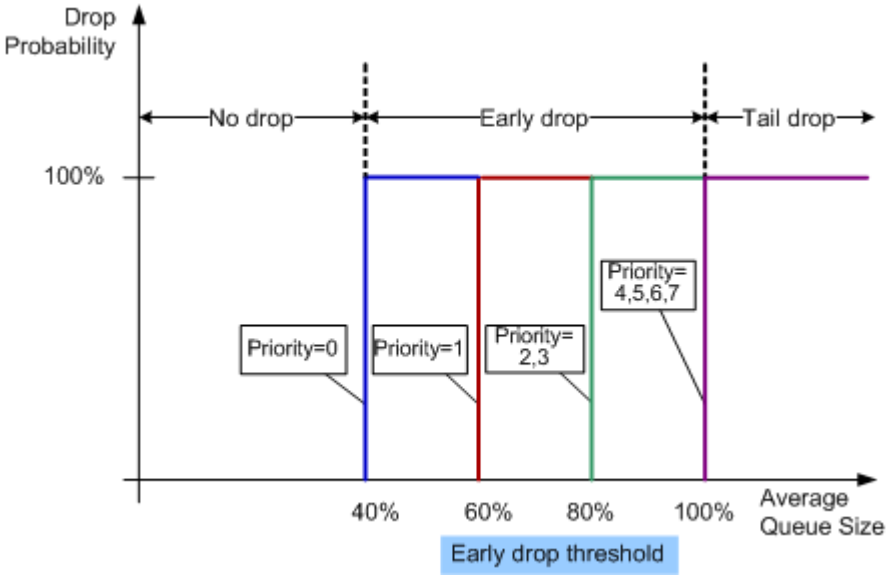
As shown in **Figure 15-29**, assume that packets with different priorities enter the same queue, the early drop thresholds are set to 40% and 60% for 0-priority and 1-priority packets, respectively, to 80% for 2-priority and 3-priority packets, and to 100% for 4-, 5-, 6-, and 7-priority packets.

- When the packets in the queue are less than 40% of the queue depth, all subsequent packets with different priorities can be enqueued.

 Copyright © Huawei Technologies Co., Ltd.

- When packets in the queue increase to or exceed 40% of the queue depth, subsequent 0-priority packets are dropped while packets with higher priorities are enqueued.
- When packets in the queue increase to or exceed 60% of the queue depth, subsequent 1-priority packets are also dropped while packets with higher priorities (except 0 and 1) are enqueued.
- When packets in the queue increase to or exceed 80% of the queue depth, subsequent 2-priority and 3-priority packets are also dropped, while packets with higher priorities (4, 5, 6, and 7) are enqueued.
- When packets in the queue increase to 100% of the queue depth, tail drop occurs. In this case, all subsequent packets are dropped.

Figure 15-29 Priority-based early drop



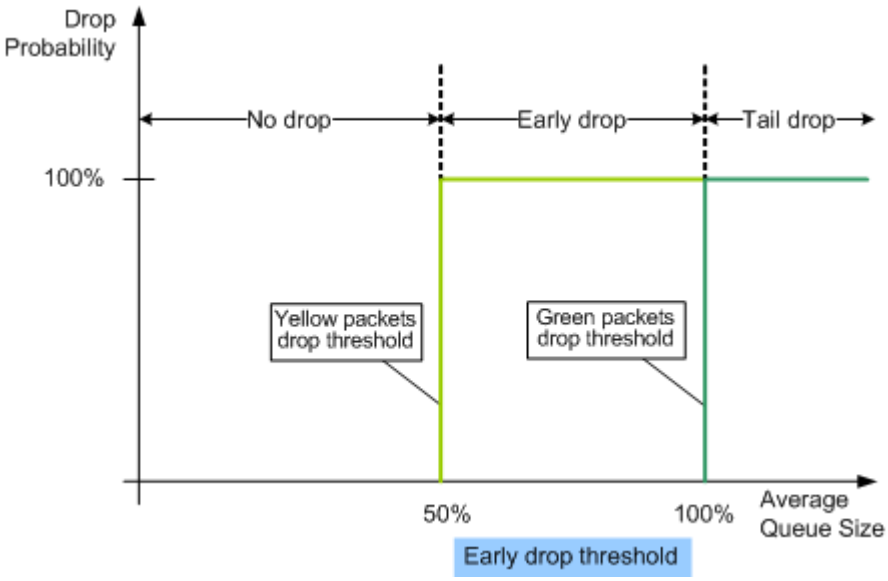
Color-based Early Drop

The system assigns different drop thresholds to different packets by marking the packets with different colors (yellow, green, or red; red packets are dropped directly) using the two rate three-color marker (trTCM) algorithm. In this way, when traffic congestion occurs on a port but the port queue is not fully filled, packets that do not exceed CIR (such packets are marked green) and packets that exceed CIR but do not exceed PIR (such packets are marked yellow) are forwarded; packets that exceed PIR (such packets are marked red) are dropped.

As shown in **Figure 15-30**, the drop threshold is 50% for yellow packets and 100% for green packets. The drop thresholds cannot be modified manually. If more than 50% of a queue is occupied, subsequent yellow packets cannot enter the queue but green packets can. When the queue is 100% occupied, green packets are also dropped.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-30 Color-based early drop



In addition, users can run the **early-drop color yellow** command to set the drop threshold for yellow packets with different priorities. Early drop for yellow packets is implemented in a similar way to priority-based early drop shown in **Figure 15-29**. The difference is that the former applies only to yellow packets with different priorities.

Color-based Early Drop - Weighted Random Early Discard (WRED)

The system assigns different drop thresholds to different packets by marking the packets with different colors (yellow, green, or red; red packets are dropped directly) using the trTCM algorithm. In this way, when traffic congestion occurs on a port but the port queue is not fully filled, packets that do not exceed CIR (such packets are marked green) and packets that exceed CIR but do not exceed PIR (such packets are marked yellow) are forwarded; packets that exceed PIR (such packets are marked red) are dropped.

The **wred-profile** command can be used to set the low drop limit (**low-limit**) and high drop limit (**high-limit**) for yellow and green packets, and the drop probability (**discard-probability**) for red packets.

The drop probability at a specific time point derives from the following formula: Packet drop probability = (Usage of the packet buffer area - Low drop limit) ÷ (High drop limit - Low drop limit) x Drop probability of the packets reaching the high drop limit.

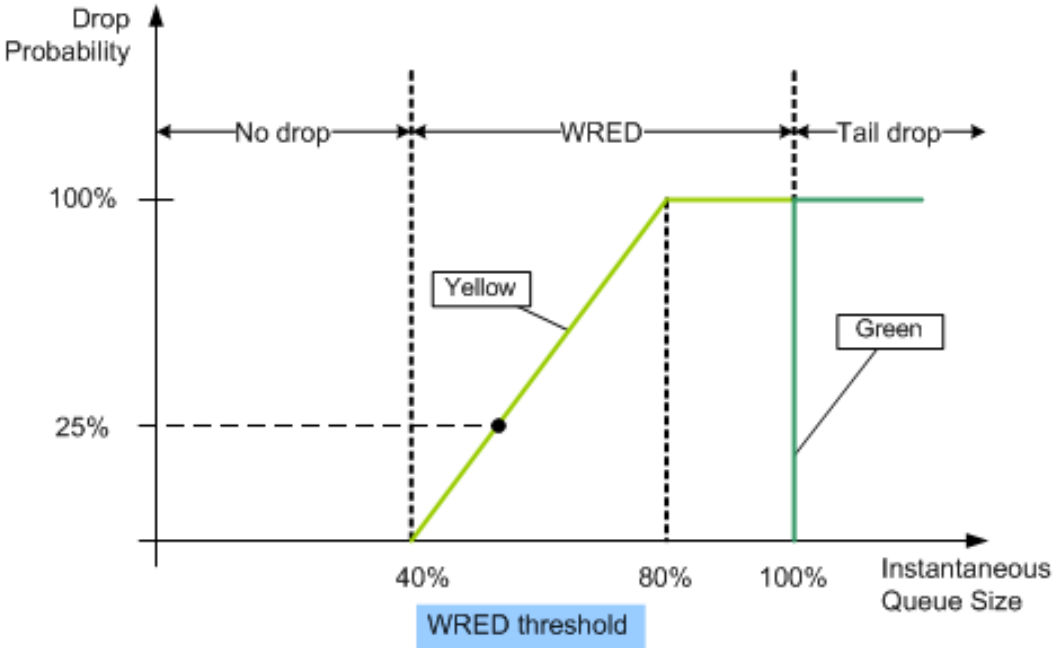
As shown in **Figure 15-31**, run the **wred-profile** command to specify the low drop limit to 40%, high drop limit to 80%, and drop probability to 100% for yellow packets, and no drop for green packets.

- When yellow packets in the queue are less than 40% of the queue depth, all subsequent yellow and green packets are enqueued.
- When yellow packets in the queue increase to or exceed 40% of the queue depth, all subsequent green packets are enqueued while subsequent yellow packets are dropped based on the drop probability specified for this time point (for example, 25%).

 Copyright © Huawei Technologies Co., Ltd.

- When yellow packets in the queue increase to or exceed 80% of the queue depth, all subsequent yellow packets are directly dropped, while all subsequent green packets are enqueued.
- When packets in the queue increase to 100% of the queue depth (the queue is fully filled), tail drop occurs. In this case, all subsequent packets are dropped.

Figure 15-31 Color-based early drop - WRED



NOTE

In the figure above, the yellow line represents the drop probability curve.

15.9.4 Configuring the Congestion Avoidance

Context

- Congestion avoidance is implemented using packet drop algorithms, which include tail drop, priority-based early drop, and color-based weighted random early detection (WRED) early drop.
- The MA5800 can differentiate the services with different priorities in the same queue. The packet priority serves as a criterion for dropping packets.
- The MA5800 can also implement early drop based on the color of packets. When congestion occurs, the yellow packets are dropped.

Procedure

- Configure priority-based early drop.
  a. Configure the early drop mode.

     Run the **early-drop mode pri-base** command to configure the priority-based early drop. After the configuration is completed, the system performs early drop according to the outer 802.1p priorities of the

 Copyright © Huawei Technologies Co., Ltd.

packets. When congestion occurs in a queue, the packets are dropped according to the early drop thresholds of the priorities.

b. Configure the early drop threshold.

Run the **early-drop** command to configure the mapping between service priorities and drop thresholds. After configuration is successful, if the packets of the specified service priority reach the threshold of the queue (the percentage of the queue depth), subsequent packets of the same service priority will be dropped instead of entering the queue.

c. Query the configured early drop threshold.

Run the **display early-drop** command to query the configured early drop threshold.

- Configure color-based early drop.
  a. Configure the early drop mode.

  Run the **early-drop mode color-base** command to configure the color-based early drop.

  According to the *CIR* and *PIR* parameters in the IP traffic profile, the system marks packets with colors. The packets within the CIR bandwidth are marked as green, and the packets between the CIR and PIR bandwidth are marked as yellow.

  After the configuration is completed, green packets are allowed to pass, yellow packets that do not exceed the bandwidth can also pass, and yellow packets that exceed the bandwidth are dropped.

  b. (Optional) Configure the early drop threshold of yellow packets.

  Run the **early-drop color yellow** command to configure the mapping between service priorities of yellow packets and drop thresholds. After the configuration is successful, if the yellow packets of the specified service priority reach the threshold of the queue (the percentage of the queue depth), subsequent yellow packets will be discarded instead of entering the queue.

  c. Query the configured early drop threshold.

  Run the **display early-drop** command to query the configured early drop threshold.

- Configure color-based WRED early drop.
  a. Configure the early drop mode.

  Run the **early-drop mode color-wred** command to configure the color-based WRED early drop.

  In this mode, a WRED profile can bind to a packet queue. Then the packet color-based early-discard can be implemented according to the parameter settings in the WRED profile

  b. Set a WRED profile.

  Run the **wred-profile** command to set the early discard threshold and discard ratio of the green or yellow packet.

  c. Bind the packet queue to the WRED profile.

 Copyright © Huawei Technologies Co., Ltd.

Run the **queue-wred** command to bind the packet queue to the WRED profile. After the WRED profile is bound, the color-based early discard is performed for packets according to the parameters in the WRED profile.

----End

Example

To configure the priority-based early drop, where,

- The early drop thresholds are set to 40% and 60% for 0-priority and 1-priority packets.
- The early drop thresholds are set to 80% for 2-priority and 3-priority packets.
- The early drop thresholds are set to 100% for 4-, 5-, 6-, and 7-priority packets.

Do as follows.

```
huawei(config)#early-drop mode pri-base
huawei(config)#early-drop cos0 40 cos1 60 cos2 80 cos3 80
{ <cr>|cos4<K>|cos5<K>|cos6<K>|cos7<K> }:
  Command:
        early-drop cos0 40 cos1 60 cos2 80 cos3 80
huawei(config)#display early-drop
{ <cr>|color<K>|mode<K>|pri-base<K>||<K> }:
  Command:
        display early-drop
  Early-drop is pri-base
  Pri-base early-drop config:
  ---------------------------------
  Priority             Threshold
  ---------------------------------
       0                  40
       1                  60
       2                  80
       3                  80
       4                 100
       5                 100
       6                 100
       7                 100
  ---------------------------------
  Color-base early-drop config:
  ---------------------------------
  Queue ID     WRED Profile Index
  ---------------------------------
       0                  -
       1                  -
       2                  -
       3                  -
       4                  -
       5                  -
       6                  -
       7                  -
  ---------------------------------
  Pri-base early-drop mapping:
  ----------------------------------
  Priority  Queue ID  Mapping Class
  ----------------------------------
       0       0           0
       1       1           0
       2       2           0
       3       3           0
       4       4           0
       5       5           0
       6       6           0
       7       7           0
  ----------------------------------
```

 Copyright © Huawei Technologies Co., Ltd.

To configure the color-based early drop, where,

- The early drop thresholds are set to 40% and 60% for 0-priority and 1-priority packets.
- The early drop thresholds are set to 80% for 2-priority and 3-priority packets.
- The early drop thresholds are set to 100% for 4-, 5-, 6-, and 7-priority packets.

Do as follows.

```
huawei(config)#early-drop mode color-base
huawei(config)#early-drop color yellow cos0 40 cos1 60 cos2 80 cos3 80 cos4  100 cos5 100 cos6 100
cos7 100
huawei(config)#display early-drop color yellow
{ <cr>||<K> }:
  Command:
        display early-drop color yellow
  Color-base early-drop config:
  ----------------------------------
  Queue ID   Yellow Packet Threshold
  ----------------------------------
        0                    40
        1                    60
        2                    80
        3                    80
        4                   100
        5                   100
        6                   100
        7                   100
  ----------------------------------
```

To configure the color-based WRED early drop, where,

- All packet queues are bound to WRED profile 0.
- Green packets are not discarded; for yellow packets, the upper discard threshold is 80, that is, all yellow packets are discarded when the usage of the packet buffer exceeds 80%.

Do as follows.

```
huawei(config)#early-drop mode color-wred
huawei(config)#wred-profile index 0 green low-limit 100 high-limit 100 discard-probability 0 yellow
low-limit 80 high-limit 80 discard-probability 100
  Command:
        wred-profile index 0 green low-limit 100 high-limit 100 discard-probab
ility 0 yellow low-limit 80 high-limit 80 discard-probability 100
                                                    huawei(config)#queue-wred queue0 0 queue1 0
queue2 0 queue3 0 queue4 0 queue5 0 queue6 0 queue7 0

Command:
        queue-wred queue0 0 queue1 0 queue2 0 queue3 0 queue4 0 queue5 0 queue
6 0 queue7 0
huawei(config)#display early-drop
{ <cr>|color<K>|mode<K>|pri-base<K>||<K> }:
  Command:
        display early-drop
  Early-drop is color-wred
  Pri-base early-drop config:
  ---------------------------------
  Priority             Threshold
  ---------------------------------
        0                    40
        1                    60
        2                    80
        3                    80
        4                   100
        5                   100
        6                   100
        7                   100
```

 Copyright © Huawei Technologies Co., Ltd.

```
---------------------------------
Color-base early-drop config:
---------------------------------
Queue ID     WRED Profile Index
---------------------------------
     0                  0
     1                  0
     2                  0
     3                  0
     4                  0
     5                  0
     6                  0
     7                  0
---------------------------------
Pri-base early-drop mapping:
-----------------------------------
Priority  Queue ID  Mapping Class
-----------------------------------
     0       0           0
     1       1           0
     2       2           0
     3       3           0
     4       4           0
     5       5           0
     6       6           0
     7       7           0
-----------------------------------
```

15.10 Congestion Management

15.10.1 Introduction to Congestion Management

Definition

If packets arrive faster than they are forwarded on a port, traffic congestion occurs on the port. Congestion management is used to manage and control traffic congestion.

Congestion management is implemented using queuing techniques for transmitting packets out of a queue.

Purpose

When traffic congestion occurs on an outbound interface, a proper queue scheduling mechanism guarantees required QoS parameters (such as bandwidth, latency, and jitter) for a certain type of packets.

15.10.2 Basic Concepts of Congestion Management

Congestion management is implemented using queuing techniques. The MA5800 supports the following queuing techniques:

- Priority queuing (PQ)
- Deficit weighted round robin (DWRR)
- PQ+DWRR

The following table explains the basic concepts of congestion management.

 Copyright © Huawei Technologies Co., Ltd.

Table 15-11 Basic concepts

Concept	Description	Remarks
PQ	In PQ, the strict priority scheduling algorithm is used. In this scheduling algorithm, eight priorities are defined. Packets with higher priorities are scheduled first.	Packets of important services are processed preferentially.
DWRR	In the Deficit WRR scheduling algorithm, packets are scheduled based on the assigned weights.	Specific QoS guarantees are provided for each queue.
PQ +DWRR	The combination of PQ+DWRR ensures that some of the packets with high priorities are scheduled first and the remaining packets are scheduled based on the specified weights.	This mechanism ensures effective transmission of services with high priorities. In the mean time, services with low priorities are scheduled in a timely manner when there is available bandwidth.
Enqueuing priority	The enqueuing priority of a packet determines the queue that the packet will enter.	There are several sources of the enqueuing priority for a packet and the priority value ranges from 0 to 7.
Queue buffering	The queue depth determines the queue's capability of processing burst packets. A larger buffering space means a better capability of processing burst packets and a lower chance to lose packets, but a larger delay in processing packets.	-

15.10.3 Implementation Principle of Congestion Management

The procedure before packets enter queues or when they are in queues is as follows:

1. Traffic classification: For details, see the **implementation principle for traffic classification**.
2. Enqueuing: Packets enter different queues based on the mapping between packet priorities and queues.
3. Scheduling: Schedules packets using a specific scheduling algorithm after the packets are enqueued.

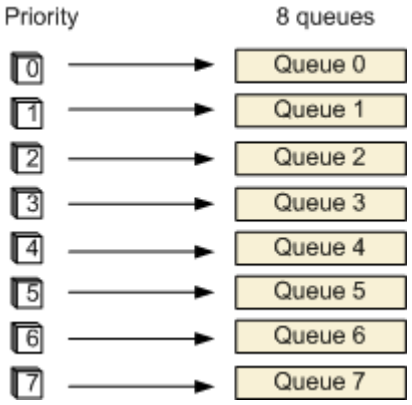
 Copyright © Huawei Technologies Co., Ltd.

Enqueuing

After priority processing, packets enter different queues based on the mapping between packet priorities and queues. By default, the mapping between packet priorities and queues is constant, as shown in **Figure 15-32**.

The larger the queue ID, the higher the forwarding priority for packets in the queue. Among all the eight queues, queue 7 has the highest priority.

Figure 15-32 Default mapping between packet priorities and queues



The system also supports a flexible mapping between packet priorities and queues. You can run the **cos-queue-map** command to map the 802.1p priority of the packet to any queue. That is, a queue can contain multiple 802.1p priorities or no 802.1p priorities.

The configuration of the mapping between packet priorities and queues takes effect globally. Generally, use the default value in the system and do not modify the parameters unless you have specific requirements.

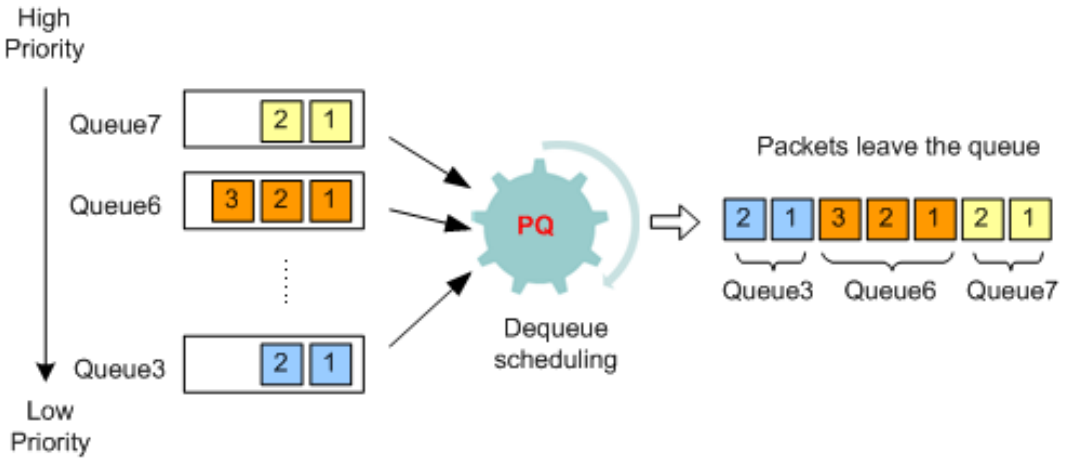
Queue Scheduling: PQ

Priority queuing (PQ) queues are classified into high-priority queues, medium-priority queues, normal-priority queues, and low-priority queues in the descending priority order. As shown in **Figure 15-33**, PQ allows the packets in a high-priority queue to exit the queue and be transmitted. After such a transmission is completed, PQ performs the same on all packets in a medium-priority queue, a normal-priority queue, and then a low-priority queue one by one.

In this way, packets in a queue with a higher priority precede packets in a queue with a lower priority and therefore are processed preferentially, even in case of congestion. This mechanism ensures that packets for critical services are processed first. Packets of non-critical services (such as email service) are processed only when the network has sufficient resources after critical services have been processed, thereby utilizing network resources efficiently.

 Copyright © Huawei Technologies Co., Ltd.

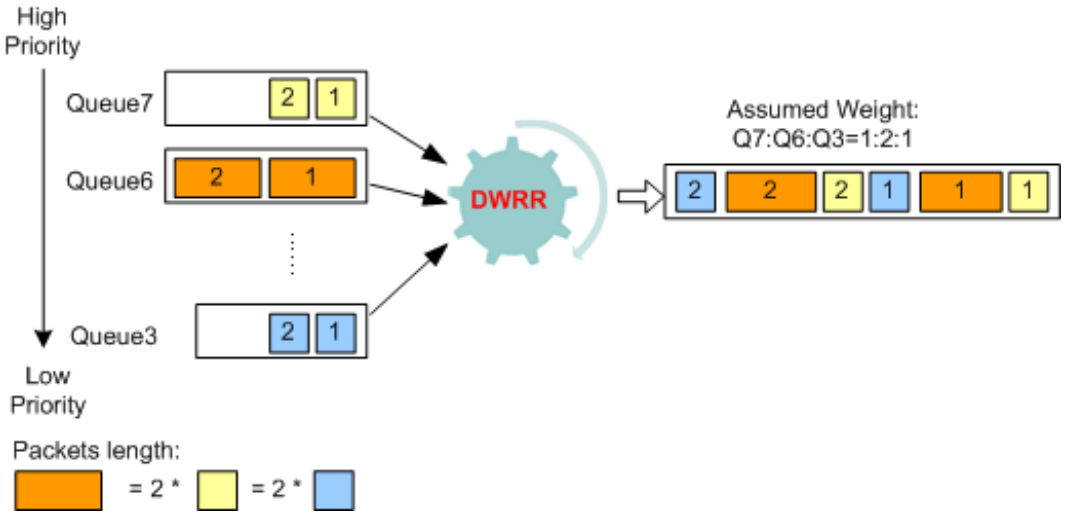
Figure 15-33 PQ scheduling mechanism



Queue Scheduling: DWRR

Deficit weighted round robin (DWRR) is a weighted queue scheduling algorithm based on packet length in the unit of byte. DWRR schedules packets by byte. In DWRR, the quantum value indicates the bytes that can be scheduled in each round, and the deficit value indicates the total bytes that can be scheduled by each queue. If the deficit value of a queue is greater than the length of the packets located at the front end of the queue, these packets are scheduled and the length of these packets is subtracted from the deficit value. Packets in a queue can be scheduled if the deficit value is greater than or equal to 0. After packets are scheduled, the deficit value becomes negative. When the deficit value is greater than or equal to 0 again, packets can be scheduled, as shown in **Figure 15-34**.

Figure 15-34 DWRR scheduling mechanism



DWRR scheduling prevents packets in low-priority queues from being starved out, that is, failing to be scheduled, and allows bandwidths to be allocated to packets

 Copyright © Huawei Technologies Co., Ltd.

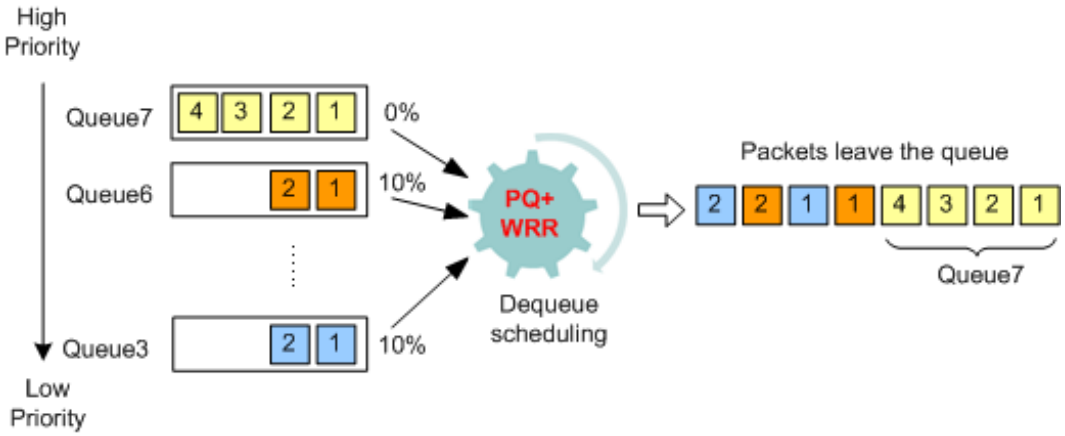
based on the weight ratio when the lengths of packets in different queues vary or change greatly.

Queue Scheduling: PQ+DWRR

PQ+DWRR is a combination of the PQ scheduling mode and DWRR scheduling mode. When the weight value of a queue is 0, the queue scheduling mode is PQ +DWRR. As shown in **Figure 15-35**, in this mode, the system schedules the queues with the weight value 0 in the PQ mode, and then schedules other queues in the DWRR mode.

This scheduling mode is flexible. The services that must be guaranteed are scheduled in PQ mode, and the services with lower priorities are scheduled in DWRR mode when there is available bandwidth. In this way, services with higher priorities are ensured and those with lower priorities can obtain bandwidth whenever there is available bandwidth.

Figure 15-35 PQ+DWRR scheduling mechanism



15.10.4 Configuring the Congestion Management

Context

Congestion management contains the following content:

- Mapping between the queue and the 802.1p priority
- Queue depth
- Queue scheduling mode

And, the ultimate effect of the congestion management is determined by all of them.

Procedure

Step 1 Configure the mapping between the queue and the 802.1p priority.

The configuration of the mapping between priorities and queues takes effect globally. By default, 802.1p priorities and queues have a one-to-one mapping

 Copyright © Huawei Technologies Co., Ltd.

relationship. Generally, use the default value in the system and do not customize the parameters unless you have specified requirements.

After the configuration, packets with different 802.1p priorities are mapped to the specified queues based on the configured mapping. This enhances the flexibility of mapping packets to queues.

1. Run the **cos-queue-map** command to configure the mapping between the 802.1p priority and the queue.
2. Run the **display cos-queue-map** command to query the mapping between the 802.1p priority and the queue.

Step 2 (Optional) Configure the queue depth.

By default, the queue depth of queue 0-7 is 13%, 6%, 13%, 13%, 12%, 12%, 25% and 6% respectively. Do not modify the value unless for special requirement.

Configure the queue depth (the queue buffer space) to re-allocate buffer space to the queues, therefore to improve the flexibility of QoS.

1. Run the **queue-buffer** command to set the buffer size percentage for packet queues of the service boards in the system.
2. Run the **display queue-buffer** command to query the buffer size percentage for queues of the service boards in the system.

Step 3 Configure the queue scheduling mode.

By default, the PQ mode is used. Properly set according to your actual requirements.

1. Run the **queue-scheduler** command to configure the queue scheduling mode.

   NOTE

   Range of *queue0-weight*: 0-100, 255, where 0 indicates that the strict PQ scheduling mode is used and 255 indicates that the queue is not used. The sum of the weight values of all queues must be 0 or 100.
2. Run the **display queue-scheduler** command to query the configuration of the queue scheduling mode.

----End

Example

Configure the queue scheduling mode to 3PQ+5DWRR. Where,

- Queues 5-7 adopt the PQ mode
- Queues 0-4 adopt the DWRR mode
- Queues 0-4 with the weights of 10, 30, 20, 20, and 20 respectively
- Other parameters use default settings

After that, packets with priority 5, 6,or 7 are scheduled in the PQ mode, and packets with other priority are scheduled in the DWRR mode.

```
huawei(config)#queue-scheduler wrr
{ queue0-weight<U><0,255> }:10
{ queue1-weight<U><0,255> }:30
{ queue2-weight<U><0,255> }:20
{ queue3-weight<U><0,255> }:20
```

 Copyright © Huawei Technologies Co., Ltd.

```
{ queue4-weight<U><0,255> }:20
{ queue5-weight<U><0,255> }:0
{ queue6-weight<U><0,255> }:0
{ <cr>|queue7-weight<U><0,255> }:0

Command:
        queue-scheduler wrr 10 30 20 20 20 0 0 0
                                                    huawei(config)#display queue-scheduler
 Queue scheduler mode : WRR
 ---------------------------------
 Queue  Scheduler Mode  WRR Weight
 ---------------------------------
    0  WRR                 10
    1  WRR                 30
    2  WRR                 20
    3  WRR                 20
    4  WRR                 20
    5  PQ                  --
    6  PQ                  --
    7  PQ                  --
 ---------------------------------
```

15.11 ACL

Using the preset access control list (ACL) policy, the system permits or refuses data packets to pass.

15.11.1 Overview of ACL

Definition

The access control list (ACL) policy defines a series of matching rules, according to which the packets to be filtered are identified. The packets identified are permitted or refused to pass according to the preset rules.

ACL-based traffic filtering is a prerequisite for quality of service (QoS). ACL together with QoS improves system security.

Benefits

Mutual access between internal networks and communication between internal and external networks are primary requirements for enterprise networks. To ensure internal network security, a security policy is required to allow unauthorized users to access specified network resources so as to control access. With the ACL, network traffic can be filtered and network access is controlled.

After being bound to a QoS operation, the ACL helps to implement the following functions:

- Limiting network traffic and improving network performance
- Ensuring safe network access
- Determining which type of communication traffic to be forwarded or blocked on a port on the network device, such as a router or switch

Controlling packets on an access port when the ACL is used for packet-filtering ACL firewall, prohibiting unauthorized users from logging in to the system and ensuring device safety

 Copyright © Huawei Technologies Co., Ltd.

15.11.2 Basic Concepts of ACL

ACL Type

The system supports IPv4 and IPv6 ACLs, that is, ACLv4 and ACLv6. ACLs include basic ACLs, advanced ACLs, link-layer ACLs, and user-defined ACLs.

Table 15-12 ACL types

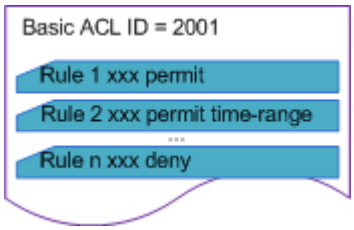
Type	Serial Number Range	Feature
Basic ACL	2000-2999	ACLv4: The rules of a basic ACL can be defined only according to the Layer 3 source IP address and the fragment field, for analyzing and processing data packets. ACLv6: The rules of a basic ACL can be defined only according to the Layer 3 source IP address and subnet prefix length, for analyzing and processing data packets.
Advanced ACL	3000-3999	Compared with a basic ACL, an advanced ACL allows for a wider scope of more accurate and flexible definition of the rules according to the following data packet information: • Source address • Destination address • IP bearer protocol types 0-255 (GRE, ICMP/ICMPv6, IP/IPv6, IPinIP, TCP, IPv6-ah, IPv6-esp, and OSPF) • TCP source port • TCP destination port • ICMP/ICMPv6 protocol type • ICMP/ICMPv6 code • Priority TOS/IP PRE/DSCP
Link-layer ACL	4000-4999	The rules of a link-layer ACL can be defined according to the following information: • Source MAC address • VLAN ID • Layer 2 protocol type • Destination MAC address • 802.1p priority

 Copyright © Huawei Technologies Co., Ltd.

Type	Serial Number Range	Feature
User-defined ACL	5000-5999	The rules of a user-defined ACL can be defined according to any 32 bytes of the first 80 bytes in a Layer 2 frame. • IPoE matching: matches packets whose Ethernet packet header is IPv4-encapsulated, including untagged IP packets, one-tagged IP packets, and two-tagged IP packets. • Non-IPoE matching: matches IPv4 packets whose Ethernet packet header is not IP-encapsulated, that is non-IPoE IPv4 packets, including untagged, one-tagged, and two-tagged non-IPoE packets, or multi-tagged packets.

Rule

Rules are the main body of an ACL. An ACL can be composed of multiple rules, as shown in the following figure.



A rule has the following characteristics:

- Each rule belongs to an absolute ACL and is of the same type as the ACL.
- Each rule corresponds to a permit or deny action.
- Each rule can be configured with a time range in which the ACL takes effect.

15.11.3 ACL Rule Matching Sequence

An ACL rule needs to be bound to a QoS policy so that it can take effect on a port. Two general ACL matching principles are provided as follows:

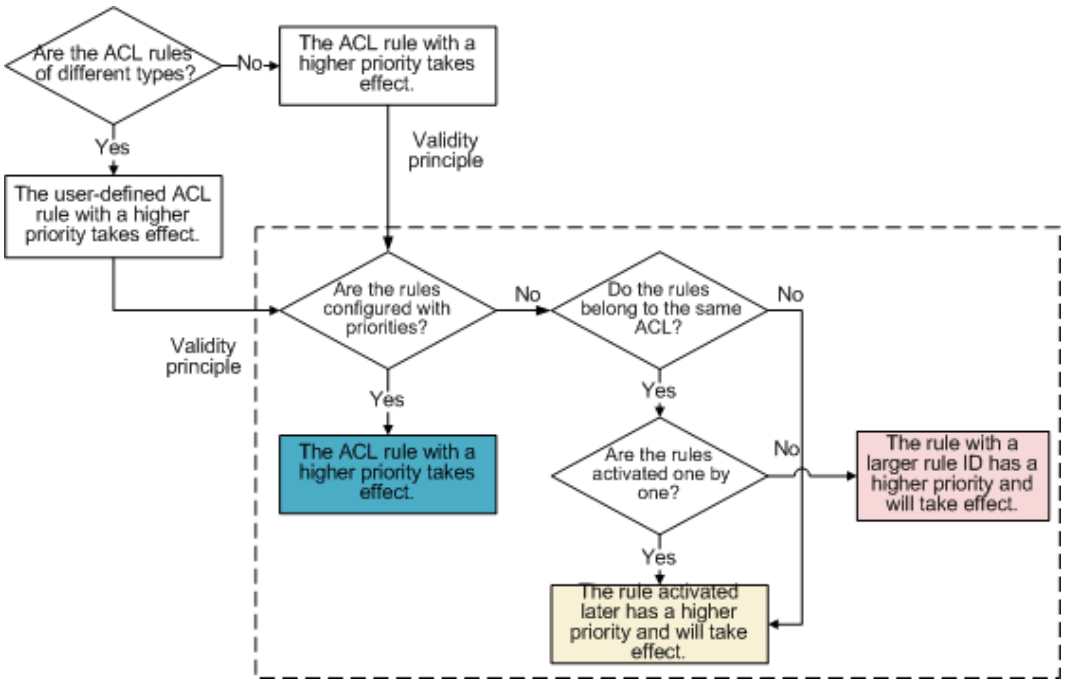
- An ACL rule is valid only when it is within the period of *time-range-name*.
- ACL rules are matched based on their priorities in a descending order; this process stops once a rule is matched.

Packets Matching Two or More ACL Rules

When packets reach a port, the system matches the packets with two or more ACL rules, as shown in **Figure 15-36**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-36 Validity of ACL rules when packets match multiple ACL rules



When packets match two or more ACL rules:

- If ACL rules are of different types, the priority of a user-defined rule is higher than that of all non-user-defined rules. If the user-defined rule is used, the other rules may be invalid. Therefore, exercise caution when using this rule.
- If ACL rules are of the same type but configured with different priorities, the ACL rule of a higher priority will be valid.
- If ACL rules are of the same type and no priority is configured for these ACL rules, do as follows:
  - For rules of the same ACL, if they are activated at the same time, the rule with a larger *rule-id* has a higher priority.
  - For rules of the same ACL, if they are activated one by one, the rule activated later has a higher priority than the one activated earlier.
  - For rules belonging to different ACLs, the rule activated later has a higher priority than the one activated earlier.

In other cases:

- When both the Layer 3 ACL (basic ACL and advanced ACL) and the Layer 2 ACL (link-layer ACL) are issued, all rules use the priority that is configured for the Layer 2 ACL rule.
- When both the IPv6-based rule and the link-layer rule exist, the link-layer rule prevails even if a higher priority is configured for the IPv6-based rule.
- Among the rules issued to the routing interface or firewall, a rule with a smaller *rule-id* has a higher priority, regardless of the activation sequence or the configured priority. The rules are used to match the packets based on **rule-id** in an ascending order. Once the rule with a smaller **rule-id** matches

 Copyright © Huawei Technologies Co., Ltd.

the packets, the rule matching stops, that is, other rules with a larger **rule-id** are invalid.

- When you run the **packet-filter** command to use an ACL and specify the **to-cpu** packet, the rules are matched based on priorities in a descending order; this process stops once a rule is matched. The matching sequence is irrelevant to whether the rule is a user-defined rule, an IPv6 rule, or a rule configured with a priority. Rules are matched based on the following principles only:
  - For rules of the same ACL, if they are activated at the same time, the rule with a larger *rule-id* has a higher priority.
  - For rules of the same ACL, if they are activated one by one, the rule activated later has a higher priority than the one activated earlier.
  - For rules belonging to different ACLs, if they are issued to the port from different ACLs, the rule activated later has a higher priority than the one activated earlier.

Packets Matching No ACL Rule

When no packet matches any rule in the ACL, the processing is as follows:

- For the ACL referenced by route interfaces or firewalls, if no packet matches any rule in the ACL, the traffic behavior will be performed, that is, the default operation is permit. You can run the **firewall default** command to configure the default operation for the firewall to **permit** or **deny**.

  For example, when you run the **firewall packet-filter** command to filter out the packets that are about to pass the firewall, and the packet does not match any rule in the ACL, the packet is allowed to pass the firewall.
- For the ACL referenced by common ports, if no packet matches any rule in the ACL, no traffic behavior will be performed.

  For example, when you run the **packet-filter** command to filter out the packets, and the packet does not match any rule in the ACL, the device will receive this packet. When you run the **traffic-limit** command to limit the traffic of packets that match the ACL rule and the packet does not match any rule in the ACL, the device only forwards the packet and does not perform the traffic limiting.

15.11.4 ACL Rule Matching Process

ACL Rule Matching Process

The system will match the inbound packets according to the defined ACL rules:

- If the packets match an ACL rule, they are performed with further QoS actions, including packet filtering, priority marking, rate limiting, traffic statistics measurement, packet redirection, and packet mirroring. After being processed using the preceding QoS actions, the packets are forwarded in the outbound direction.
- If the packets do not match an ACL rule, the packets are directly forwarded.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

However, when the ACL rule is applied to the packet filtering firewall, the packets that do no match the ACL rule are processed according to the default firewall policy (run the **firewall default { permit | deny }** command to configure the default firewall policy). If the firewall default policy is deny, the packets will be discarded.

The following table lists the ACL-based QoS policies and relevant functions and configuration methods.

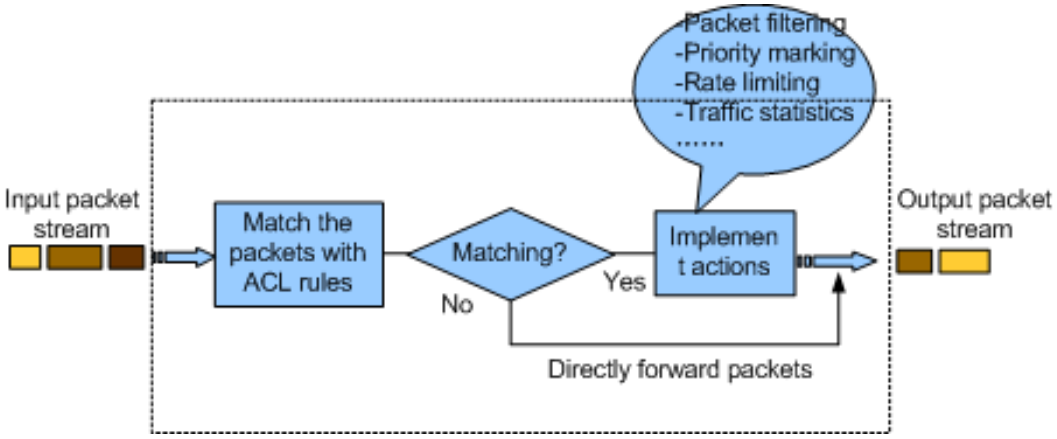
ACL-based QoS Policy	Function	Configuration Method
Packet filtering	Determines whether to discard packets or allow the packets to pass according to the matching result of the ACL rule (deny or permit).	Configuration command: **packet-filter**
Priority marking	Marks priorities for packets that match an ACL rule on a specified port or port list by type, such as ToS, DSCP, and 802.1p.	Configuration command: **traffic-priority**
Rate limiting	Limits the rate of packets that match an ACL rule on a specified port or port list. If traffic on a port exceeds the limit, the excessive packets will be discarded based on the trTCM algorithm or be configured with a new DSCP priority.	Configuration command: **traffic-limit**
Traffic statistics	Measures the packets that match an ACL rule on a specified port or port list in terms of packet number and byte.	Configuration command: **traffic-statistic**
Packet redirection	Redirects the packets that match an ACL rule on a specified port or port list to a new forwarding destination port (the original port is not used to forward packets that match an ACL rule).	Configuration command: **traffic-redirect**

 Copyright © Huawei Technologies Co., Ltd.

ACL-based QoS Policy	Function	Configuration Method
Packet mirroring	Mirrors the packets that match an ACL rule on a specified port and copies the packets to other ports. If packets received or transmitted through a port need to be monitored during device maintenance or fault diagnosis, run the command to mirror packets to be monitored to other ports. **NOTE** Based on industry experience, the mirroring feature may involve obtaining personal data of users and the content of users' communications (the product does not save, parse, or process such information) for the purpose of safeguarding network operation and protecting services. Huawei alone is unable to collect or save the personal data of users and the content of users' communications. It is suggested that you activate the interception-related functions based on the applicable laws and regulations in terms of purpose and scope of usage. You are obligated to take considerable measures to ensure that the personal data of users and the content of users' communications are fully protected when the personal data and the content are being used and saved.	Configuration command: **traffic-mirror**

Figure 15-37 illustrates ACL-based packet filtering and processing.

Figure 15-37 ACL-based packet filtering and processing



 Copyright © Huawei Technologies Co., Ltd.

15.11.5 Matching Principle for the User-defined ACL Rule

The user-defined ACL rule is matched using any 32 bytes of the first 80 bytes of a Layer 2 data frame. The following describes the format of the Layer 2 data frame.

Format of the Layer 2 Data Frame

The following uses a Layer 2 IPv4 data frame as an example.

Figure 15-38 First 64 bytes of an IPv4 data frame

Byte	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	AA	AA	AA	AA	AA	AA	BB	BB	BB	BB	BB	BB	CC	CC	CC	CC
	DD	DD	EE	FF	GG	GG	HH	HH	II	II	JJ	KK	LL	LL	MM	MM
	MM	MM	NN	NN	NN	NN	OO	OO	PP	PP	QQ	QQ	QQ	QQ	RR	RR
	RR	RR	SS	TT	UU	UU	VV	VV	VV	VV	VV	VV	VV	VV	VV	VV

Byte	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	AA	AA	AA	AA	AA	AA	BB	BB	BB	BB	BB	BB	CC	CC	CC	CC
	DD	DD	EE	FF	GG	GG	HH	HH	II	II	JJ	KK	LL	LL	MM	MM
	MM	MM	NN	NN	NN	NN	OO	OO	PP	PP	QQ	QQ	QQ	QQ	RR	RR
	RR	RR	SS	TT	UU	UU	VV	VV	VV	VV	VV	VV	VV	VV	VV	VV

shows the sequence of the first 64 bytes of the Layer 2 IPv4 data frame, in which:

- Different letters indicate different field values. For example, A refers to the destination MAC address and C refers to the Ethernet type+VLAN tag.
- The first letter of each letter group indicates the offset value of the field. For example, A indicates that the offset value is 0 and C indicates that the offset value is 12.

Table 15-13 lists the meaning and offset value of each letter.

Table 15-13 Meanings of letters and their offset values

Letter & Meaning	Offset Value	Letter & Meaning	Offset Value	Letter & Meaning	Offset Value
A: destination MAC address	0	I: flags	24	Q: serial number	42
B: source MAC address	6	J: time to live	26	R: acknowledgment field	46

 Copyright © Huawei Technologies Co., Ltd.

Letter & Meaning	Offset Value	Letter & Meaning	Offset Value	Letter & Meaning	Offset Value
C: Ethernet type +VLAN tag	12	K: protocol ID ("6" represents TCP and "17" represents UDP)	27	S: IP header length and reserved bit	50
D: protocol type	16	L: IP checksum	28	T: reserved bit and flags bit	51
E: IP version number	18	M: source IP address	30	U: window size	52
F: service type	19	N: destination IP address	34	V: Others	54
G: length of the IP packet	20	O: TCP source port	38	-	-
H: ID	22	P: TCP destination port	40	-	-

NOTE

The offset value of each field is the offset value in data frame ETH II+VLAN tag.

Matching Principle

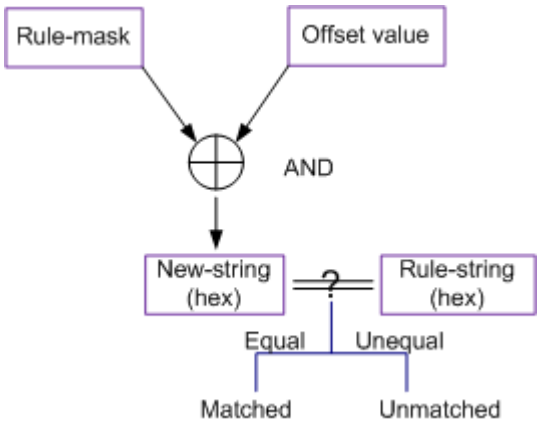
The user-defined ACL rule involves the following key concepts:

- rule-string: indicates the character string of the user-defined ACL rule. The character string is in hexadecimal notation. The number of characters in the string must be an even number ranging from 2 to 8.
- rule-mask: indicates the mask of the user-defined ACL rule. It is used to perform the logical AND operation with specified fields in the data packets.
- offset: indicates the offset. With the header of the packet as the reference point, it specifies a certain field in the Layer 2 data frame.

With the user-defined ACL rule, users can use rule-mask and offset to extract any 32 bytes of the first 80 bytes of a Layer 2 data frame and compare them with the rule-string so as to find the matched data frame. **Figure 15-39** shows the matching principle.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-39 Matching principle for the user-defined ACL rule



15.11.6 Configuring Traffic Management Based on ACL Rules

The ACL can be used to implement flexible traffic classification according to user requirements. After traffic classification based on ACL rules is completed, you can perform QoS for the traffic streams.

15.11.6.1 Configuration Differences Between IPv4 ACLs and IPv6 ACLs

This topic describes differences regarding to configuration between IPv4 ACLs and IPv6 ACLs. It is recommended that you know well about how to configure IPv4 ACLs and then configure IPv6 ACLs based on their differences.

Configuration Differences

- The configuration differences between IPv4 ACLs and IPv6 ACLs are as follows:
  - IPv6 and IPv4 have different IP address formats and packet formats, so the **ipv6** parameter must be specified for configuring IPv6 basic ACLs and advanced ACLs. Use the **ipv6** parameter to choose between IPv4 ACLs and IPv6 ACLs.
  - IPv4 and IPv6 have the same link-layer packet encapsulation format, so configurations do not differentiate IPv6 link-layer ACLs and IPv4 link-layer ACLs.
  - Users define packets matching ACLs based on the packet type. IPv4 and IPv6 have the same packet command for user-defined ACLs, so configurations do not differentiate IPv6 user-defined ACLs and IPv4 user-defined ACLs. When user-defined ACLs are used for filtering packets, the protocol type of the packets must be the same as the protocol type of the ACL rules. If they are different, filtering may encounter errors.

15.11.6.2 Configuring an ACL Rule

 Copyright © Huawei Technologies Co., Ltd.

Context

ACLs include:

- Basic ACL
- Advanced ACL
- Link-layer ACL
- User-defined ACL

For features of each type of ACL, see "**ACL Type**."

Because the ACL is flexible in use, Huawei provides the following suggestions on its configuration:

- It is recommended that you define a general rule, such as permit any or deny any, in each ACL, so that each packet has a matching traffic rule that determines to forward or filter the unspecified packet.
- The activated ACL rules share the hardware resources with the protocol modules, such as the Dynamic Host Configuration Protocol (DHCP) module and Internet Protocol over ATM (IPoA) module. In this case, the hardware resources are limited and may be insufficient. To prevent the failure to enable other service functions due to insufficient hardware resources, it is recommended that you enable the protocol module first and then activate ACL rules in the data configuration. If you fail to enable a protocol module, perform the following steps:
  a. Check whether ACL rules occupy too many resources.
  b. If ACL rules occupy too many resources, deactivate or delete the unimportant or temporarily unused ACL configurations, and then configure and enable the protocol module.

ACL-based applications can be classified to:

- Traffic management based on a single ACL rule (base mode)
- Traffic management based on a group of ACL rules (enhance mode)

For these 2 applications, the command **qos-acl mode** can be executed to switch the base and enhance modes. In these 2 modes, the method and commands of adding ACL sub-rules are the same. The configuration mode cannot be switched when:

- QoS is configured in the system.
- Remote TCP packet header obtaining and traffic burst detection are configured in the system.

Procedure

- Configure a basic ACL rule.

  It is applicable to the scenario where the device needs to match packets by source IP address.

  a. (Optional) Set a time range.

     Run the **time-range** command to create a time range, which can be used when an ACL rule is created.

 Copyright © Huawei Technologies Co., Ltd.

b. Create a basic ACL.

Run the **acl** command to create a basic ACL, and then enter the ACL mode. The serial number of a basic ACL ranges from 2000 to 2999.

c. Configure a basic ACL rule.

In the acl-basic mode, run the **rule** command to create a basic ACL rule. The parameters are as follows:

- *rule-id*: indicates the ACL rule ID. To create an ACL rule with a specified ID, use this parameter.
- **permit**: indicates the keyword for allowing the data packets that meet related conditions to pass.
- **deny**: indicates the keyword for discarding the data packets that meet related conditions.
- **time-range**: indicates the keyword of the time range during which the ACL rule will take effect.

- Configure an advanced ACL rule.

It is applicable to the scenario where the device needs to match data packets by source IP address, destination IP address, type of protocol running over IP, and protocol feature, such as TCP source port, TCP destination port, and ICMP type of the data packets.

a. (Optional) Set a time range.

Run the **time-range** command to create a time range, which can be used when an ACL rule is created.

b. Create an advanced ACL.

Run the **acl** command to create an advanced ACL, and then enter the acl-adv mode. The serial number of an advanced ACL ranges from 3000 to 3999.

c. Configure an advanced ACL rule.

In the acl-adv mode, run the **rule** command to create an ACL rule. The parameters are as follows:

- *rule-id*: indicates the ACL rule ID. To create an ACL rule with a specified ID, use this parameter.
- **permit**: indicates the keyword for allowing the data packets that meet related conditions to pass.
- **deny**: indicates the keyword for discarding the data packets that meet related conditions.
- **time-range**: indicates the keyword of the time range during which the ACL rule will take effect.

- Configure a link-layer ACL rule.

It is applicable to the scenario where the device needs to match packets by link-layer information such as source MAC address, source VLAN ID, Layer 2 protocol type, and destination MAC address.

 Copyright © Huawei Technologies Co., Ltd.

a. (Optional) Set a time range.

Run the **time-range** command to create a time range, which can be used when an ACL rule is created.

b. Create a link-layer ACL.

Run the **acl** command to create a link-layer ACL, and then enter the acl-link mode. The serial number of a link-layer ACL ranges from 4000 to 4999.

c. Configure a link-layer ACL rule.

In the acl-link mode, run the **rule** command to create a link-layer ACL rule. The parameters are as follows:

- *rule-id*: indicates the ACL rule ID. To create an ACL rule with a specified ID, use this parameter.
- **permit**: indicates the keyword for allowing the data packets that meet related conditions to pass.
- **deny**: indicates the keyword for discarding the data packets that meet related conditions.
- **time-range**: indicates the keyword of the time range during which the ACL rule will take effect.

- Configure a user-defined ACL rule.

It is applicable to the scenario where the device needs to match packets by any 32 bytes of the first 80 bytes of a Layer 2 data frame.

Configuring a user-defined ACL requires a deep understanding of the Layer 2 data frame structure. Be sure to make a data plan according to the format of the Layer 2 data frame. Refer to "**15.11.5 Matching Principle for the User-defined ACL Rule**" for details.

a. (Optional) Set a time range.

Run the **time-range** command to create a time range, which can be used when an ACL rule is created.

b. Create a user-defined ACL.

Run the **acl** command to create a user-defined ACL, and then enter the acl-user mode. The serial number of a user-defined ACL ranges from 5000 to 5999.

c. Configure a user-defined ACL rule.

In the acl-user mode, run the **rule** command to create an ACL rule. The parameters are as follows:

- *rule-id*: indicates the ACL rule ID. To create an ACL rule with a specified ID, use this parameter.
- **permit**: indicates the keyword for allowing the data packets that meet related conditions to pass.
- **deny**: indicates the keyword for discarding the data packets that meet related conditions.

 Copyright © Huawei Technologies Co., Ltd.

- *rule-string*: indicates the character string of the user-defined ACL rule. The character string is in hexadecimal notation. The number of characters in the string must be an even number.
- *rule-mask*: indicates the mask of the user-defined ACL rule. It is a positive mask, used to perform the AND operation with the data packets for extracting the information from the data packets.
- *offset*: indicates the offset. With the header of the packet as the reference point, it specifies the byte from which the AND operation begins. Together with the rule mask, it extracts a character string from the packets.
- **ipoe**: indicates that the Ethernet packet header encapsulates an IP packet, including untagged, single-tagged, and double-tagged IP packets.
- **non-ipoe**: indicates that the Ethernet packet header encapsulates a non-IP packet, including the untagged, single-tagged, and double-tagged non-IP packets, or multi-tagged packets.
- **time-range**: indicates the keyword of the time range during which the ACL rule will take effect.

- Perform traffic management based on ACL group.

  When you need to apply an ACL group on a port, switch QoS ACL mode to enhance mode.

  a. Configure the QoS ACL mode.

     Run the **qos-acl mode** command to set enhance mode.

  b. Create ACL and configure sub-rules. For details, see the preceding steps.

  c. Apply packet filtering and traffic limitation on ACL group-matched packets of a specified port.

----End

Example

To configure port 0/2/0 on the MA5800 to receive only the packets from address 10.2.2.2 from 00:00 to 12:00 on Fridays, and to discard the packets from other addresses, do as follows:

```
huawei(config)#time-range time1 00:00 to 12:00 fri
huawei(config)#acl 2000
huawei(config-acl-basic-2000)#rule deny time-range time1
huawei(config-acl-basic-2000)#rule permit source 10.2.2.2 0.0.0.0 time-range time1
huawei(config-acl-basic-2000)#quit
huawei(config)#packet-filter inbound ip-group 2000 port 0/2/0
huawei(config)#save
```

Assuming that the service board of the MA5800 resides in slot 1 and belongs to a VLAN, and the IP address of the VLAN Layer 3 interface is 10.10.10.101, to prohibit the ICMP (such as ping) and telnet operations from the user side to the VLAN interface on the device, do as follows:

```
huawei(config)#acl 3001
huawei(config-acl-adv-3001)rule 1 deny icmp destination 10.10.10.101 0
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-acl-adv-3001)rule 2 deny tcp destination 10.10.10.101 0 destination-port eq telnet
huawei(config-acl-adv-3001)quit
huawei(config)#packet-filter inbound ip-group 3001 rule 1 port 0/2/0
huawei(config)#packet-filter inbound ip-group 3001 rule 2 port 0/2/0
huawei(config)#save
```

NOTE

Configuration can be performed based on ACL group to achieve the same function.

```
huawei(config)#qos-acl mode enhance
huawei(config)#acl 3001
huawei(config-acl-adv-3001)rule 1 deny icmp destination 10.10.10.101 0
huawei(config-acl-adv-3001)rule 2 deny tcp destination 10.10.10.101 0 destination-port eq telnet
huawei(config-acl-adv-3001)quit
huawei(config)#packet-filter inbound ip-group 3001 rule port 0/2/0
huawei(config)#save
```

Pay attention to the following during the configuration:

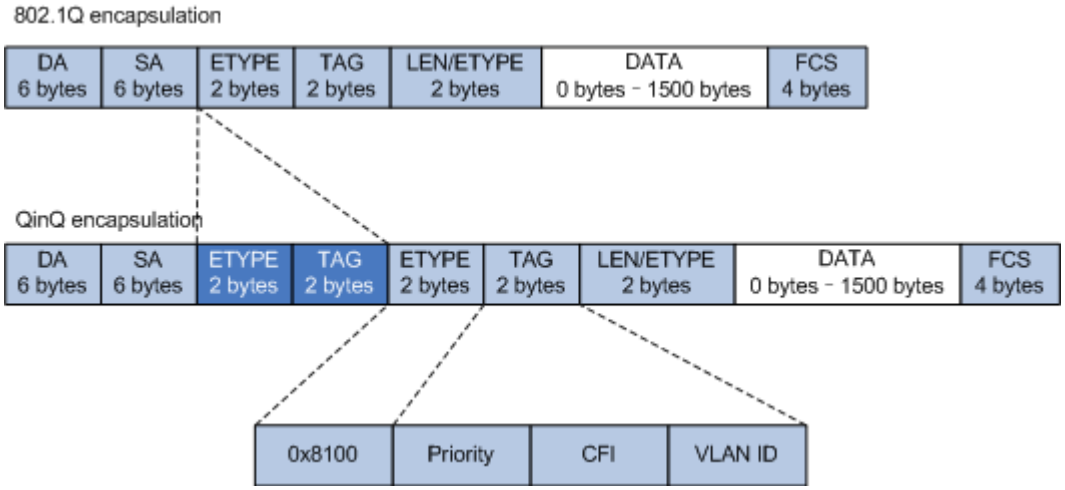
- If packets are filtered based on ACL group and parameter **ruleid** is not specified, all rules defined in ACL group are used to filter packets on a port.
- When a rule is added to or deleted from an ACL group, the action bound to this group is delivered or deleted. You do not need to configure this rule.

To create a link-layer ACL rule that allows data packets with protocol type 0x8863 (pppoe-control message), VLAN ID 12, CoS 1, source MAC address 00e0-fc00-2222, and destination MAC address 00e0-fc11-4141 to pass, do as follows:

```
huawei(config)#acl 4001
huawei(config-acl-link-4001)rule 1 permit type 0x8863 cos 1 source 12  00e0-fc00-2222 0000-0000-0000 destination 00e0-fc11-4141 0000-0000-0000
huawei(config-acl-link-4001)quit
huawei(config)#save
```

Assuming that the packet sent from port 0/2/0 to the MA5800 is the QinQ packet containing two VLAN tags, to change the CoS priority in the outer VLAN tag (VLAN ID: 10) to 5, do as follows:

Figure 15-40 QinQ packet format




```
huawei(config)#acl 5001
huawei(config-acl-user-5001)#rule 1 permit 8100 ffff 16
```

 Copyright © Huawei Technologies Co., Ltd.

NOTE

The type value of a QinQ packet varies with vendors. Huawei adopts the default 0x8100. As shown in **Figure 15-40**, the offset of this type value should be 16 bytes.

```
huawei(config-acl-user-5001)#rule 10 permit 0a ff 19
huawei(config-acl-user-5001)#quit
```

NOTE

"19" indicates the ADN operation after an offset of 19 bytes with the header of the packet as the base. "0a" refers to the value of the inner tag field of the QinQ packet. In this example, the second byte of the inner tag field is a part of the VLAN ID, which is exactly the value of the inner VLAN ID (VLAN 10).

```
huawei(config)#traffic-priority inbound user-group 5001 cos 5 port 0/2/0
```

15.11.6.3 Configuring ACL Matching for PPPoE Packets

This topic describes the format of PPPoE packets and how to match user-defined ACLs for various PPPoE packets.

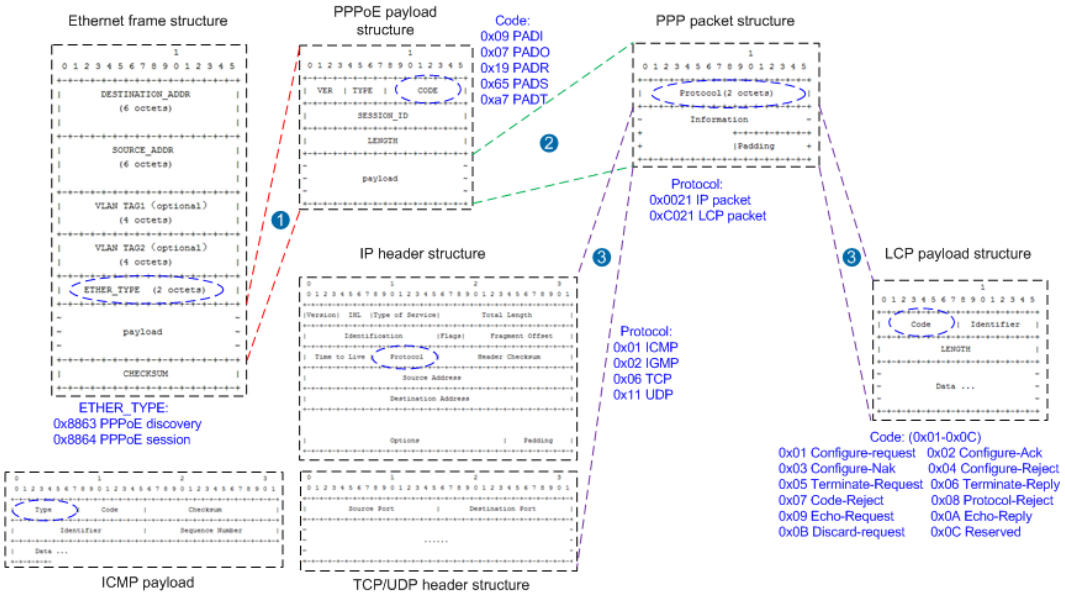
Prerequisites

You are familiar with the **15.11.5 Matching Principle for the User-defined ACL Rule**.

Context

Figure 15-41 shows the PPPoE packet format and offset value of each field.

Figure 15-41 PPPoE packet format and offset value of each field



Basic knowledge about PPPoE packets:

- In PPPoE-based data transmission, data packets are carried in the payload of ETH packets for transmission.

 Copyright © Huawei Technologies Co., Ltd.

- The whole PPPoE-based transmission process has two stages: discovery (ETHER_TYPE = 0x8863) and session (ETHER_TYPE = 0x8864).
- In the discovery stage, there are four types of packets: PADI (code = 0x09), PADO (code = 0x07), PADR (code = 0x19), and PADS (code = 0x65).
- In the session stage, data is transmitted through PPP and PPP data is carried in the payload of PPPoE packets.
- PPP data frames can carry IP data packets (protocol = 0x0021) or LCP data packets (protocol = 0xC021).
- Based on the value of the protocol field in the IP header of IP data packets, the data transmission mode can be determined. The data transmission modes include:
  - 0x01 ICMP
  - 0x02 IGMP
  - 0x06 TCP
  - 0x11 UDP
- Based on the value (value range: 0x01-0x0C) of the code field of the LCP data packets, the packet encapsulation type can be determined.

Configuration Guideline

1. Determine the stage status of PPPoE packets, discovery or session.
2. Determine the protocol used to carry PPP packets, IP or LCP.
3. Determine the transmission mode (UDP, IGMP, or others) or encapsulation type of data packets.
4. Determine the offset value of each field.

Procedure

- **Match ACLs for PADI packets with a specified source MAC address and a single VLAN tag.**

  As shown in **Figure 15-41**, PADI packets contain the following information:
  - Packets are in the discovery stage.
  - Since the value of the ETHER_TYPE field of ETH packets is 0x8863, and packets carry only one single VLAN tag, the offset value is 16 (6 + 6 + 4 = 16).
  - Since the value of the code field of PPPoE packets is 0x09, the offset value is 19 (16 + 2 + 1 = 19).

  Based on the **15.11.5 Matching Principle for the User-defined ACL Rule**, the corresponding ACLs can be determined.

  Example: Run the following commands to match ACLs for PADI packets with the source MAC address of 0x00e0-fc00-4455 and a single VLAN tag:

  ```
  huawei(config)#acl 5001
  huawei((config-acl-user-5001))#rule permit 00e0fc00 FFFFFFFF 6 4455 FFFF 10 8863 FFFF 16 09 FF 19 non-ipoe
  ```

- **Match ACLs for PPPoE echo-request packets with two VLAN tags.**

  As shown in **Figure 15-41**, PPPoE echo-request packets contain the following information:

 Copyright © Huawei Technologies Co., Ltd.

- Packets are in the session stage and LCP is used for packet transmission.
- Since the value of the ETHER_TYPE field of ETH packets is 0x8864, and packets carry two VLAN tags, the offset value is 20 (6 + 6 + 4 + 4 = 20).
- Since the value of the protocol field of PPP packets is 0xC021, the offset value is 28 (20 + 2 + 6 = 28).
- Since the value of the code field of LCP packets is 0x09, the offset value is 30 (28 + 2 = 30).

Based on the **15.11.5 Matching Principle for the User-defined ACL Rule**, the corresponding ACLs can be determined.

Example: Run the following commands to match ACLs for PPPoE echo-request packets with two VLAN tags:

```
huawei(config)#acl 5001
huawei((config-acl-user-5001))#rule permit 8864 FFFF 20 C021 FFFF 28 09 FF 30 non-ipoe
```

- **Match ACLs for packets with a specified inner VLAN ID and a specified UDP source port number.**

  As shown in **Figure 15-41**, the packets contain the following information:

  - Packets are in the session stage and IP UDP is used for packet transmission.
  - Since the value of the ETHER_TYPE field of ETH packets is 0x8864, and packets carry two VLAN tags, the offset value is 20 (6 + 6 + 4 + 4 = 20).
  - Since the value of the protocol field of PPP packets is 0x0021, the offset value is 28 (20 + 2 + 6 = 28).
  - Since the value of the protocol field in the IP header is 0x11, the offset value is 39 (28 + 2 + 9 = 39).
  - Based on the UDP source port number, the offset value is 50 (39 + 1 + 10 = 50).

  Based on the **15.11.5 Matching Principle for the User-defined ACL Rule**, the corresponding ACLs can be determined.

  Example: Run the following commands to match ACLs for PPPoE packets whose inner VLAN ID is 291 (0x123) and UDP source port number is 94 (0x5e):

  ```
  huawei(config)#acl 5001
  huawei((config-acl-user-5001))#rule permit 0123 0FFF 16 8864 FFFF 20 0021 FFFF 28 11 FF 39 5E FF 50 non-ipoe
  ```

- **Match ACLs for PPPoE-based PING (echo-request) packets with specified inner and outer VLAN IDs.**

  As shown in **Figure 15-41**, the packets contain the following information:

  - Packets are in the session stage and IP ICMP is used for packet transmission.
  - Since the value of the ETHER_TYPE field of ETH packets is 0x8864, and packets carry two VLAN tags, the offset value is 20 (6 + 6 + 4 + 4 = 20).
  - Since the value of the protocol field of PPP packets is 0x0021, the offset value is 28 (20 + 2 + 6 = 28).
  - Since the value of the protocol field in the IP header is 0x01, the offset value is 39 (28 + 2 + 9 = 39).

 Copyright © Huawei Technologies Co., Ltd.

– Since the value of the type field in the ICMP header is 0x08, the offset value is 50 (39 + 1 + 10 = 50).

Based on the **15.11.5 Matching Principle for the User-defined ACL Rule**, the corresponding ACLs can be determined.

Example: Run the following commands to match ACLs for PPPoE-based PING (echo-request) packets whose outer and inner VLAN IDs are 256 (0x100) and 291 (0x123) respectively.

```
huawei(config)#acl 5001
huawei((config-acl-user-5001))#rule permit 0100 0FFF 12 0123 0FFF 16 8864 FFFF 20 0021 FFFF 28 01 FF 39 08 FF 50 non-ipoe
```

----End

15.11.6.4 Controlling the Traffic Matching an ACL Rule

This topic describes how to control the traffic matching an ACL rule on a specified port, and process the traffic that exceeds the limit, such as adding the DSCP tag or dropping the packet directly.

Prerequisite

The **ACL and the rule of the ACL are configured**, and the port for traffic limit is working in the normal state.

Context

- The traffic statistics are only effective for the permit rules of an ACL.
- The limited traffic must be an integer multiple of 64 Kbit/s.

Procedure

Step 1 Run the **traffic-limit** command to control the traffic matching an ACL rule on a specified port.

Use the **target-rate** parameter to set the fixed maximum rate of the port, or use CAR parameters to set a rate for trTCM-based ports. The two rates cannot be set at a time. Run this command to set the action to be taken when the traffic received on the port exceeds the limited value. Two options are available:

- **drop**: Drop the traffic that exceeds the limited value.
- **remark-dscp** *value*: To set the DSCP priority for the traffic that exceeds the limited value, use this parameter.

Step 2 Run the **display qos-info traffic-limit port** command to query the traffic limit information on the specified port.

----End

Example

To limit the traffic that matches ACL 2001 received on port 0/2/0 to 512 Kbit/s, and add the DSCP priority tag (af1) to packets that exceed the limit, do as follows:

```
huawei(config)#traffic-limit inbound ip-group 2001 512 exceed remark-dscp af1 port 0/2/0
//"af1" represents a dscp type: Assured Forwarding 1 service (10).
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#display qos-info traffic-limit port 0/2/0
traffic-limit:
port 0/2/0:
 Inbound:
   Matches: Acl 2001 rule 5     running
     Target rate: 512 Kbps
     Exceed action: remark-dscp af1
```

15.11.6.5 Adding a Priority Tag to the Traffic Matching an ACL Rule

This topic describes how to add a priority tag to the traffic matching an ACL rule on a specified port so that the traffic can obtain the service that matches the specified priority. The priority tag type can be ToS, DSCP, or 802.1p.

Prerequisite

The **ACL and the rule of the ACL are configured**, and the port for traffic limit is working in the normal state.

Context

- The traffic statistics are only valid to permit rules of an ACL.
- The ToS and the DSCP priorities are mutually exclusive. Therefore, they cannot be configured at the same time.

Procedure

Step 1 Run the **traffic-priority** command to add a priority tag to the traffic matching an ACL rule on a specified port.

Step 2 Run the **display qos-info traffic-priority port** command to query the configured priority.

----End

Example

To add a priority tag to the traffic that matches ACL 2001 received on port 0/2/1, and the DSCP priority and local priority of the traffic are 10 (af1) and 0 respectively, do as follows:

```
huawei(config)#traffic-priority inbound ip-group 2001 dscp af1 local-precedence 0 port 0/2/1
huawei(config)#display qos-info traffic-priority port 0/2/1

traffic-priority:
port 0/2/1:
 Inbound:
   Matches: Acl 2001 rule 5  running
     Priority action: dscp af1 local-precedence 0
```

15.11.6.6 Enabling the Statistics Collection of the Traffic Matching an ACL Rule

This topic describes how to enable the statistics collection of the traffic matching an ACL rule, analyzing and monitoring the traffic.

 Copyright © Huawei Technologies Co., Ltd.

Prerequisite

The **ACL and the rule of the ACL are configured**, and the port for traffic statistics is working in the normal state.

Procedure

Step 1 Run the **traffic-statistic** command to enable the statistics collection of the traffic matching an ACL rule on a specified port.

Step 2 Run the **display qos-info traffic-statistic port** command to query the statistics information about the traffic matching an ACL rule on a specified port.

----End

Example

To enable the statistics collection of the traffic that matches ACL 2001 received on port 0/9/0, do as follows:

```
huawei(config)#traffic-statistic inbound ip-group 2001 port 0/9/0
huawei(config)#display qos-info traffic-statistic port 0/9/0

traffic-statistic:
port 0/9/0:
 Inbound:
   Matches: Acl 2001 rule 5     running
     0 packet
```

15.11.6.7 Enabling the Mirroring of the Traffic Matching an ACL Rule

This topic describes how to mirror the traffic matching an ACL rule on a port to a specified port. Mirroring does not affect packet receipt and transmission on the mirroring source port. You can monitor the traffic of the mirroring source port by analyzing the traffic that passes the mirroring destination port.

Prerequisite

The **ACL and the rule of the ACL are configured**, and the port for traffic mirroring is working in the normal state.

NOTE

Based on industry experience, the mirroring feature may involve obtaining personal data of users and the content of users' communications (the product does not save, parse, or process such information) for the purpose of safeguarding network operation and protecting services. Huawei alone is unable to collect or save the personal data of users and the content of users' communications. It is suggested that you activate the interception-related functions based on the applicable laws and regulations in terms of purpose and scope of usage. You are obligated to take considerable measures to ensure that the personal data of users and the content of users' communications are fully protected when the personal data and the content are being used and saved.

Context

- The traffic statistics are only valid to permit rules of an ACL.

 Copyright © Huawei Technologies Co., Ltd.

- The destination mirroring port cannot be an aggregation port.
- The system supports only one mirroring destination port.

Procedure

Step 1 Run the **traffic-mirror** command to enable the mirroring of the traffic matching an ACL rule on a specified port.

Step 2 Run the **display qos-info traffic-mirror port** command to query the mirroring information about the traffic matching an ACL rule on a specified port.

----End

Example

To mirror the traffic that matches ACL 2001 received on port 0/2/1 to port 0/9/0, do as follows:

```
huawei(config)#traffic-mirror inbound ip-group 2001 port 0/2/1 to port 0/9/0
huawei(config)#display qos-info traffic-mirror port 0/2/1

traffic-mirror:
port 0/2/1:
 Inbound:
   Matches: Acl 2001 rule 5     running
   Mirror to: port 0/9/0
```

15.11.6.8 Enabling the Redirection of the Traffic Matching an ACL Rule

This topic describes how to redirect the traffic matching an ACL rule on a specified port. After this operation is executed successfully, the original port does not forward the traffic matching the ACL rule, but the specified port forwards the traffic.

Prerequisites

The **ACL and the rule of the ACL are configured**, and the port for redirection is working in the normal state.

Context

- The traffic statistics are only valid to permit rules of an ACL.
- Currently, the service ports support only redirection of the traffic matching the ACL rule to upstream ports. The upstream ports support only redirection of the traffic matching the ACL rule to ports on the board of the same type.

Procedure

Step 1 Run the **traffic-redirect** command to redirect the traffic matching an ACL rule on a specified port.

Step 2 Run the **display qos-info traffic-redirect port** command to query the redirection information about the traffic matching an ACL rule on a specified port.

----End

 Copyright © Huawei Technologies Co., Ltd.

Example

To redirect the traffic that matches ACL 2001 received on port 0/9/0 to port 0/9/1, do as follows:

```
huawei(config)#traffic-redirect inbound ip-group 2001 port 0/9/0 to port 0/9/1
huawei(config)#display qos-info traffic-redirect port 0/9/0
traffic-redirect:
port 0/9/0:
 Inbound:
   Matches: Acl 2001 rule 5     running
    Redirected to: port 0/9/1
```

15.12 ACLv6

This topic describes the aspects unique to access control list for IP version 6 (ACLv6), and the differences between ACLv6 and ACLv4. For details about the specifications and principles of ACLv6, see **15.11 ACL**.

15.12.1 Comparison Between ACLv6 and ACLv4

On the MA5800, the application and the configuration processes of ACLv6 are the same as those of ACLv4, except the following differences regarding specifications and commands:

- ACLv4 supports segmented packets, while ACLv6 does not.
- IPv6 and IPv4 have different IP address formats and packet formats, so the **ipv6** parameter must be specified when you are configuring basic ACLs and advanced ACLs for IPv6. When the **ipv6** parameter is specified, the system supports both ACLv4 and ACLv6.
- Regarding user-defined ACLs, packets are matched based on the packet type defined by users. Because IPv4 and IPv6 have the same command format for specifying the packet type in user-defined ACLs, configurations do not differentiate between IPv6 user-defined ACLs and IPv4 user-defined ACLs. When user-defined ACLs are used for filtering packets, however, the protocol type of the packets must be matched against the protocol type specified in the ACL rules.

15.13 HQoS

Traditional quality of service (QoS) schedules traffic based on port. Hierarchical quality of service (HQoS) not only controls port traffic at a finer service granularity but also schedules traffic based on service priorities.

15.13.1 Overview of HQoS

Definition

Hierarchical quality of service (HQoS) is a technology that uses a multi-level scheduling mechanism to guarantee the bandwidth of various services for multiple users. It usually applies to Open Access networks.

The following HQoS scheduling models are available:

 Copyright © Huawei Technologies Co., Ltd.

- The first level of HQoS guarantees the bandwidth of services for a user, and the second level guarantees the bandwidth of all users that use the same types of services.
- The first level of HQoS guarantees the bandwidth of services for a user, and the second level guarantees the bandwidth of all services for the user.

Background

In most scenarios, traditional QoS identifies service types on a network and provides services at different levels. As users and service types increase, traditional QoS faces the following challenges:

- Traditional QoS schedules traffic based on port bandwidth. Traffic management is therefore sensitive to the service level rather than users, which applies to traffic at the network core side rather than traffic at the service access side.
- Traditional QoS is unable to uniformly manage or hierarchically schedule various services or multiple users.

Purpose

HQoS controls traffic in a user-specific and service-specific manner, and therefore guarantees bandwidth of various services for multiple users. In addition, it provides the committed information rate (CIR) and peak information rate (PIR) of various services for each user.

Benefits

Benefits to carriers

Unlike traditional QoS which schedules traffic based on a port, HQoS implements QoS on a port at a finer granularity (users and service flows). Therefore, HQoS enables a carrier to guarantee QoS for enterprises and contracted users, provide guaranteed bandwidths and service packages for more users, and achieve higher profitability.

Benefits to users

HQoS ensures that the bandwidth designated for a user is not affected by other users.

15.13.2 Open Access of HQoS

Hierarchical quality of service (HQoS) mainly applies to Open Access networks. This topic describes the Open Access network model for better understanding of HQoS.

Definition and Benefits

Open Access provides a network business model that separates the physical bearer network from the service network. The infrastructure of an Open Access network, including passive infrastructure (optical fibers, equipment room premises, and cables) and active network devices, is built by a nation or an operator authorized

 Copyright © Huawei Technologies Co., Ltd.

by the nation. Retail service providers (RSPs) directly lease bandwidth on the infrastructure network to provide service packages to end-users.

In the traditional model, an operator builds and operates its own network and delivers services to end-users. Unlike the traditional model, Open Access builds a layered network over which separate RSPs deliver their services.

Open Access brings the following benefits:

- Maximizes the freedom of choice for end users. End-users have more services to choose from and can even switch from one service provider to another without changing their home terminals (such as their ONTs).
- Lowers investment risks for RSPs. The business model of Open Access greatly shortens the cycle of return on investment (ROI). The traditional business model of operators, who usually have monopoly over their networks, requires an ROI cycle of 8-10 years. The Open Access business model shortens the ROI cycle to 1-2 years. Hence Open Access lowers the investment entrance level and risks for RSPs, and promotes competition and innovation.
- Opens up a wider arena for RSPs. RSPs no longer need to build the infrastructure network and are able to focus on innovation and competition of services and contents.

Open Access Modes

In a broad sense, there are two Open Access modes: the physical open access mode (LLU) and the bit stream open access mode.

- LLU Access Mode: This is a layer-one physical open access mode. In this mode, RSPs lease duct resources such as copper loops and optical fibers. One line cannot carry the services of multiple RSPs. To subscribe to services of different content providers (CPs), a user needs to apply for respective lines.
- Bit Stream Access Mode: In the bit stream access mode, RSPs are separated from the infrastructure network. RSPs purchase bandwidth on the infrastructure network and provide service packages to end-users.

Wholesale is a special bit stream access mode. RSPs lease physical terminals, and one terminal belongs to one RSP. Wholesale allows for simpler terminal management.

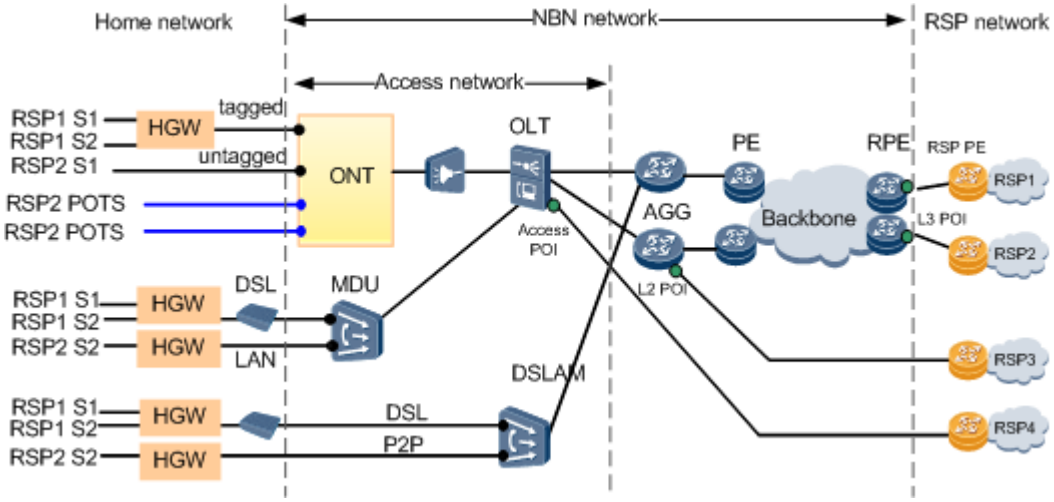
The bit stream access mode presents multiple requirements and challenges to FTTx devices. Therefore, the following part of this document will focus on the bit stream access mode.

End-to-end Open Access network

Figure 15-42 shows an end-to-end Open Access network (Bit Stream Access Mode).

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-42 End-to-end Open Access network



Concept	Description
NBN	short for national broadband network. An NBN is usually constructed using the Open Access network model.
POI	short for point of interconnection. It is a point where a carrier network and retail service provider (RSP) network interconnects. Based on network layers, there are access POI, Layer 2 POI, and Layer 3 POI.

An end-to-end Open Access network consists of the RSP networks at both ends and the NBN open channel network in between. The NBN open channel network includes the ONT, OLT, and aggregation devices such as the user-end provider edge (UPE) and access aggregation gateway (AGG).

An RSP needs to purchase physical ports and logical bandwidth to access the NBN open channel network. The physical ports are the openings at both ends of the open channel.

- User network interfaces (UNIs). They are usually the ports on user-side devices, such as the ETH/POTS/Wi-Fi ports on an ONT, xDSL/LAN ports on a multi-dwelling unit (MDU), and xDSL/P2P ports on a digital subscriber line access multiplexer (DSLAM). UNIs connect upper-layer devices to user terminals or RSPs' home gateways.
- External network-to-network interfaces (ENNIs), also called point of interconnection (POI), are the interconnection points between an operator network and an RSP network. POIs can be the upstream GE or 10GE ports on an OLT, aggregation device, or backbone network device.

Because RSPs will share one physical network, the RSPs purchase logical bandwidth on the network and use the bandwidth as service channels to provision services to end-users. The logical bandwidth includes the bandwidth on the UNIs for access users and the aggregation bandwidth on the ENNIs. The logical bandwidth is expressed in the unit of bit/s.

Openness and fairness are two key points of a bit stream Open Access network. Openness means that an ONT on user side can be shared by multiple RSPs for

 Copyright © Huawei Technologies Co., Ltd.

provisioning services. Also, on network side the ENNIs can provide various openings to meet different network requirements. Fairness means that users of the same RSP can be isolated from each other, and RSPs can be isolated from each other, through HQoS.

15.13.3 Basic Concepts of HQoS

Concept	Description
HQoS user	A basic unit of bandwidth guarantee and traffic scheduling. Alternatively, a type of user service can also be a basic unit.
HQoS user group	A group of HQoS users. The committed information rate (CIR) of HQoS users is guaranteed by scheduling an HQoS user group.
Retail service provider (RSP)	Unlike carriers, RSPs focus on service content rather than infrastructure network construction. They lease bandwidth to quickly provision user services.
Open Access	A network construction mode that separates the physical bearer network from the service network. RSPs directly lease bandwidth on an infrastructure network where Open Access has been deployed to quickly provide service packages for customers.
CAR group	A combination of service flows for unified QoS control, with configurable parameters such as CIR and peak information rate (PIR). Generally, a CAR group is specified for limiting the rate of triple-play services of a specific household user.

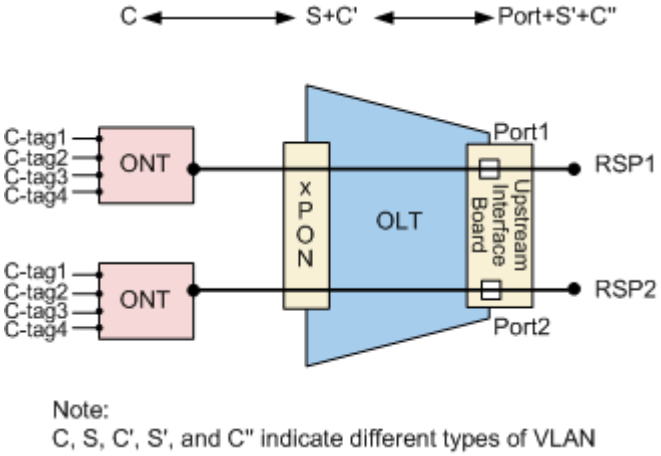
15.13.4 HQoS Service Model (Based on Port+VLAN)

Figure 15-43 shows an HQoS service processing model based on the Port+VLAN. A Layer 2 leased line is used, between the ingress port on the optical network terminal (ONT) and the uplink port of the optical line terminal (OLT). On the line, services are forwarded based on S-VLAN+C-VLAN tags.

- The ONT attaches C-VLAN tags to services based on a port or service type to identify users.
- After receiving the services, the OLT translates the C-VLAN tags into C-VLAN' tags based on PON port information and adds S-VLAN tags to the services based on service types. The services are identified based on their S-VLAN+C-VLAN' tags.
- After services are forwarded to a port on the upstream interface board that is interconnected to an RSP network, the upstream interface board translates the S-VLAN+C-VLAN' tags into S-VLAN' +C-VLAN'' tags based on the RSP planning requirements configured on the port.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-43 HQoS service model based on Port+VLAN



The Layer 2 leased line can be considered as designated to an HQoS user with the following QoS requirements:

1. CIR can be ensured when network congestion occurs, and PIR can be ensured when the network is idle.
2. The services of the user occupy bandwidth based on service priorities. Specifically, services with a higher priority can preferentially occupy CIR and PIR resources.

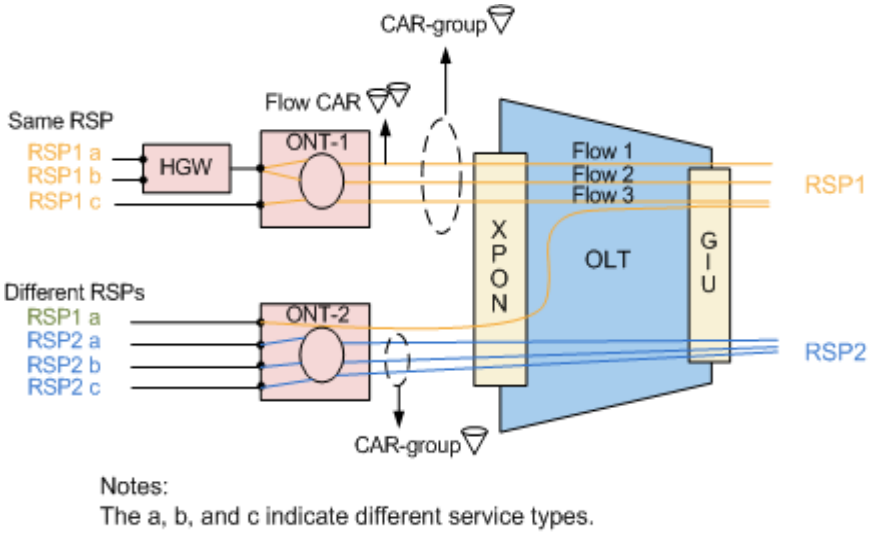
15.13.5 HQoS Service Model (Based on a CAR Group)

Mode 1

- Installed at a user's home, an optical network terminal (ONT) can run services provided by one retail service provider or RSP (such as ONT-1 shown in the figure), or services provided by multiple RSPs (such as ONT-2 shown in the figure).
- On the ONT, triple play services are provisioned and each type of service is transmitted through one GEM port. On the optical line terminal (OLT), each type of service is mapped into one service flow.
- There is no limit on the VLAN tag translation mode. An S-VLAN+C-VLAN tagging mode, however, is recommended (the S-VLAN tag indicates an RSP while the C-VLAN tag indicates a service type).

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-44 HQoS service model based on a CAR group



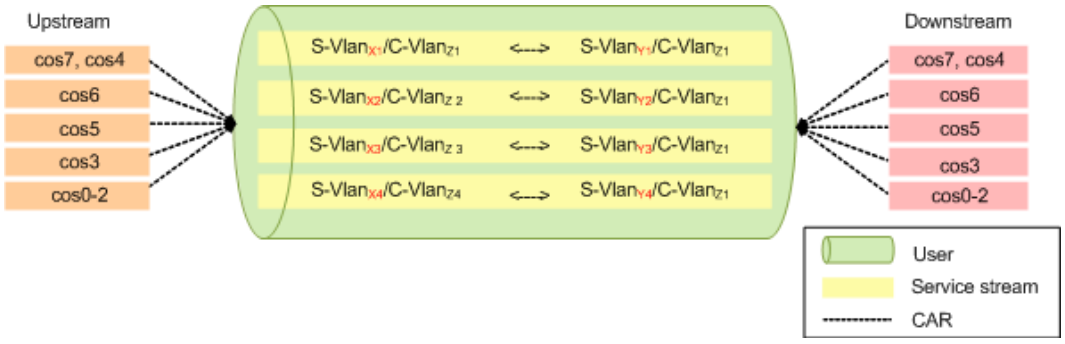
This HQoS service model can address the following QoS requirements:

1. For each type of service, CIR can be ensured when network congestion occurs, and PIR can be ensured when the network is idle.
2. When services are provided by only one RSP, the total user bandwidth can be limited.
3. When services are provided by multiple RSPs, the total lease bandwidth of each RSP can be limited.

Mode 2

For RSP users leasing pipes, services can be flexibly deployed by configuring multi-level CAR groups on the OLT. As shown in **Figure 15-45**, RSP users leasing pipes group service flows for level-1 rate limitation. Every group of service flows then can be grouped by priority for level-2 rate limitation (priority group CAR for service flows).

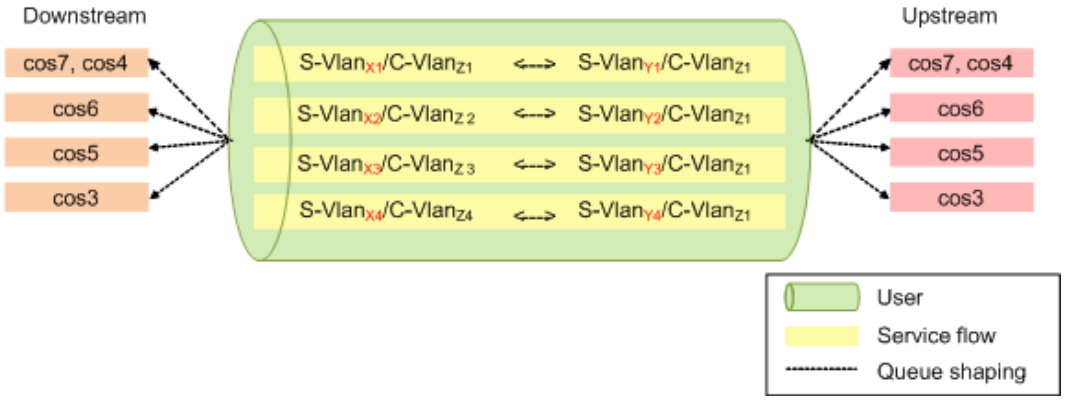
Figure 15-45 HQoS service model based on a CAR group



 Copyright © Huawei Technologies Co., Ltd.

As shown in **Figure 15-46**, subscriber shaping can be performed on the pipe for RSP users as the level-1 rate limitation and then queue shaping by priority as the level-2 rate limitation.

Figure 15-46 HQoS service model based on a CAR group

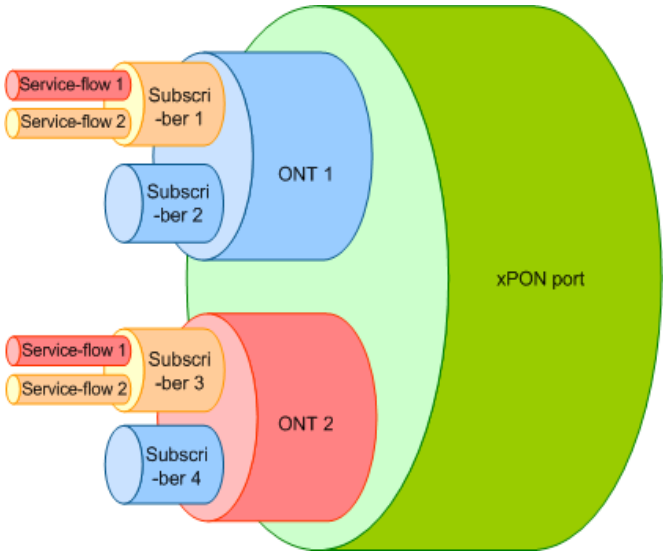


15.13.6 HQoS Service Model (xPON Board)

Figure 15-47 shows the HQoS service model of an xPON board.

- One xPON port connects to multiple ONTs.
- One ONT connects to multiple users.
- One user is provisioned with multiple services.

Figure 15-47 HQoS service model (xPON board)



As shown in the figure, this model supports the specifications for HQoS:

- Rate limitation on every user and every service

 Copyright © Huawei Technologies Co., Ltd.

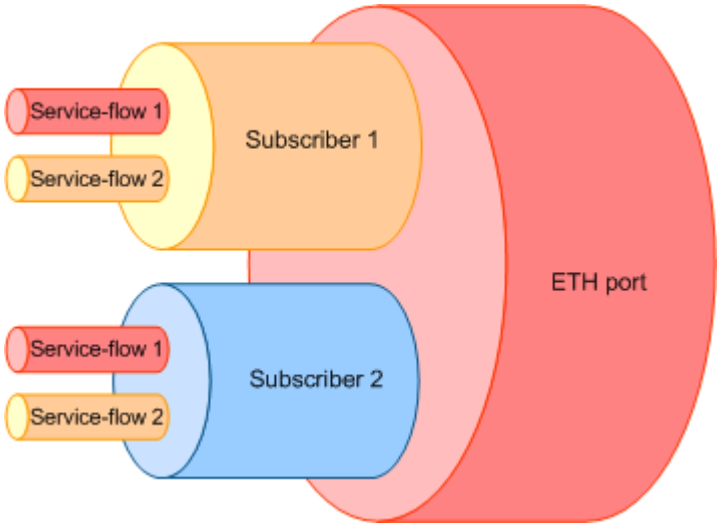
- Bandwidth scheduling between multiple services of every user
- Rate limitation on every user
- Bandwidth scheduling on every ONT
- Rate limitation on every ONT
- Scheduling between multiple ONTs connected to every xPON port

15.13.7 HQoS Service Model (P2P Board)

Figure 15-48 shows the HQoS service model of a P2P board.

- An Ethernet port connects to multiple users.
- One user is provisioned with multiple services.

Figure 15-48 HQoS service model (P2P board)



As shown in the figure, this model supports the specifications for HQoS:

- Rate limitation on every user and every service
- Packet scheduling between multiple services of every user
- Rate limitation on every user
- Packet scheduling between multiple users connected to every Ethernet port

15.13.8 Implementation Principle of HQoS

15.13.8.1 Implementation Principle of HQoS based on Port+VLAN

Implementation Model of HQoS based on Port+VLAN

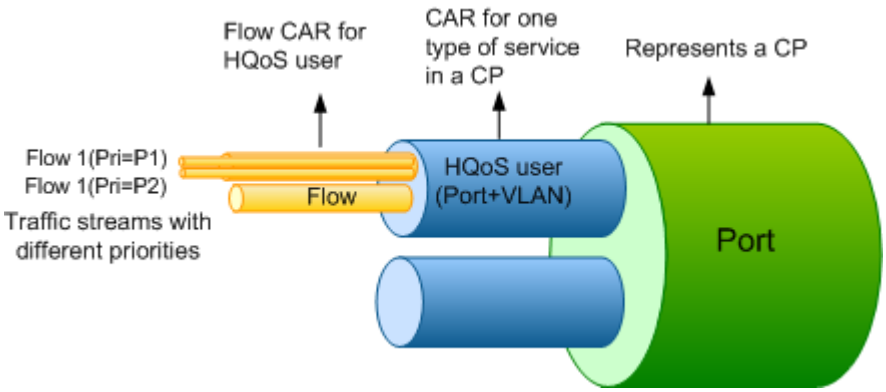
Figure 15-49 shows an implementation model of HQoS based on Port+VLAN. In the model, each HQoS user is mapped to one Layer 2 service flow.

The core of the implementation principle of HQoS based on Port+VLAN is a two-level service rate limitation mechanism on the upstream interface board. The following describes the mechanism:

 Copyright © Huawei Technologies Co., Ltd.

1. The committed access rate (CAR) is limited for an HQoS user, and the service packets are marked with colors based on committed information rate (CIR) and peak information rate (PIR). If priority-based CAR is enabled (with CAR thresholds configured), packets with a higher priority will be marked green and transmitted.
2. Color-based CAR is performed on an HQoS user group (a type of service provided by an RSP and identified by port+VLAN tags). In this manner, the bandwidth for a type of service can be ensured among the total leased RSP bandwidth, HQoS users can be isolated from each other, the CIR can be ensured for the services of HQoS users, and the PIR can also be ensured if bandwidth resources are sufficient.

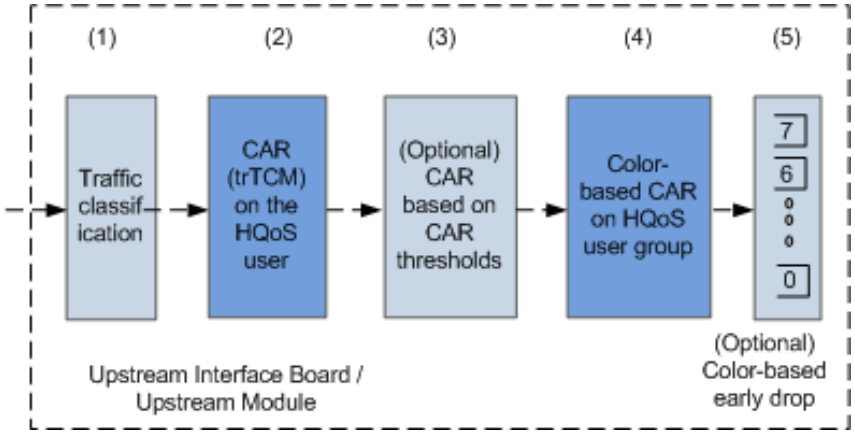
Figure 15-49 Implementation model of HQoS based on Port+VLAN



Implementation Solution

Figure 15-50 shows the implementation solution of HQoS based on port+VLAN on an upstream board.

Figure 15-50 Procedure for implementing HQoS based on port+VLAN



The following describes how HQoS is implemented using the upstream interface board:

 Copyright © Huawei Technologies Co., Ltd.

1. The upstream board performs traffic classification on received packets.

   On an optical line terminal (OLT), each HQoS user is mapped to one Layer 2 service flow. When a service packet from the uplink direction enters the OLT, the upstream board on the OLT performs traffic classification on it based on related packet parameters and determines the HQoS user to which the service packet belongs.

   - On the network-to-network interface (NNI) side (or an uplink port of the upstream board), traffic classification can be performed based on port+S-VLAN+C-VLAN' tags.
   - On the user-to-network interface (UNI) side, traffic classification can be performed based on S-VLAN+C-VLAN' tags.

   For detailed principles of traffic classification, see **Implementation Principle of Traffic Classification**.

2. The upstream board performs CAR on the HQoS user (flow CAR) using a two rate three color marker (trTCM) algorithm.

   The traffic profile bound to the Layer 2 service flow (that is, the HQoS user) is used for CAR. Specifically, a packet is marked with a color based on the CIR and PIR.

   - If the packet rate is lower than or equal to the CIR, the upstream board will mark the packet green (and will transmit it).
   - If the packet rate is higher than the CIR, but is lower than or equal to the PIR, the upstream board will mark the packet yellow (and will transmit it).
   - If the packet rate is higher than the PIR, the upstream board will directly drop the packet.

   For detailed principles of the trTCM algorithm, see **Dual-Token Bucket (trTCM) Principle**.

3. (Optional) The upstream board performs flow CAR based on CAR thresholds, and marks user packets with colors based on packet priorities (configurable through the **car-threshold** command).

   The upstream board marks the packets of the HQoS user with colors based on the CIR and PIR settings. (If an enhanced trTCM algorithm is used, thresholds for 4 priorities will be supported.) A packet will be marked green if the packet rate is lower than the CIR, or yellow if the packet rate is between the CIR and the PIR, or will be directly dropped if the packet rate is higher than the PIR. When marking packets with colors, the upstream board also differentiates packet priorities. Packets with a higher priority can preferentially use the CIR and PIR bandwidth specified for the HQoS user.

   For detailed principles of packet color marking based on CAR thresholds, see **Enhanced trTCM: Marking Packet with Color Based on CAR Thresholds**.

4. The upstream board performs CAR on HQoS user groups (group CAR) based on packet colors (configurable through the **car-port *portid* vlan** command).

   As shown in **Figure 15-51**, CAR is performed on yellow packets and green packets that have been identified in step 2. Specifically,

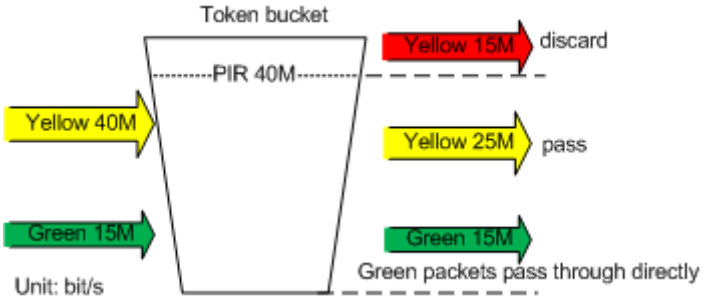
   a. Green packets will be transmitted.

   b. All yellow packets can be transmitted if their rate is within the remaining bandwidth (Remaining bandwidth = PIR - Rate of green packets).

 Copyright © Huawei Technologies Co., Ltd.

c. If the rate of yellow packets exceeds the remaining bandwidth, some yellow packets can be transmitted using the remaining bandwidth while the excessive yellow packets will be dropped.

Figure 15-51 CAR on yellow packets and green packets



5. (Optional) The upstream board performs color-based early drop on an egress port.

   On the egress port, the upstream board puts all HQoS users into the same priority queue. The queue deploys a strict priority-based scheduling policy, and color-based early drop will be performed based on the queue. When the rate of yellow packets reaches the specific drop threshold, yellow packets will be dropped to ensure low latency of green packets. Note that the drop threshold is 50% for yellow packets and 100% for green packets. The two drop thresholds cannot be modified manually.

   To put HQoS users into one priority queue, the following configuration methods can be used:

   - In fiber to the building (FTTB) scenarios where both S-VLAN and double C-VLAN tags are attached to packets, the outer S-VLAN tags of HQoS users can be set to a specific priority so that all HQoS users can be in the same queue based on the priority.
   - In fiber to home (FTTH) scenarios where packets are single-tagged, all priorities of single tags can be mapped into one port queue.

   For details on color-based early drop, see **Color-based Early Drop**.

15.13.8.2 Implementation Principle of CAR-Group-based HQoS

Implementation Model of CAR-Group-based HQoS

A CAR group is a combination of traffic streams for unified QoS control, with configurable parameters such as CIR and PIR. One of its typical applications is for multiple services (IPTV, Internet access, and voice services) of home users. Using the CAR group, QoS based on the home user instead of based on each service is implemented.

Figure 15-52 illustrates the model of CAR-group-based HQoS supported by the MA5800.

HQoS based on a CAR group implements two-level CAR for services:

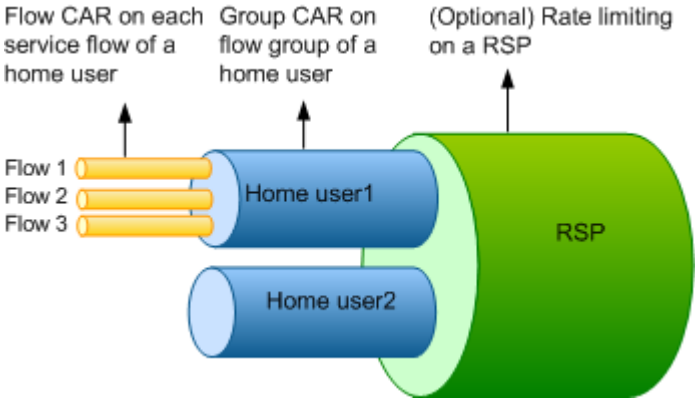
1. The first level performs flow CAR on various service flows of a home user.
2. The second level performs group CAR on flow groups of home users.

 Copyright © Huawei Technologies Co., Ltd.

If required, a level of CAR can be performed specifically for retail service providers (RSPs).

For example, the two-level CAR can achieve the following settings: limiting a user's Internet access rate, multicast service rate, and voice service rate to 2 Mbit/s, 4 Mbit/s, and 128 Kbit/s respectively, and at the same time limiting the total bandwidth of the user to 5 Mbit/s.

Figure 15-52 Implementation model of CAR-group-based HQoS

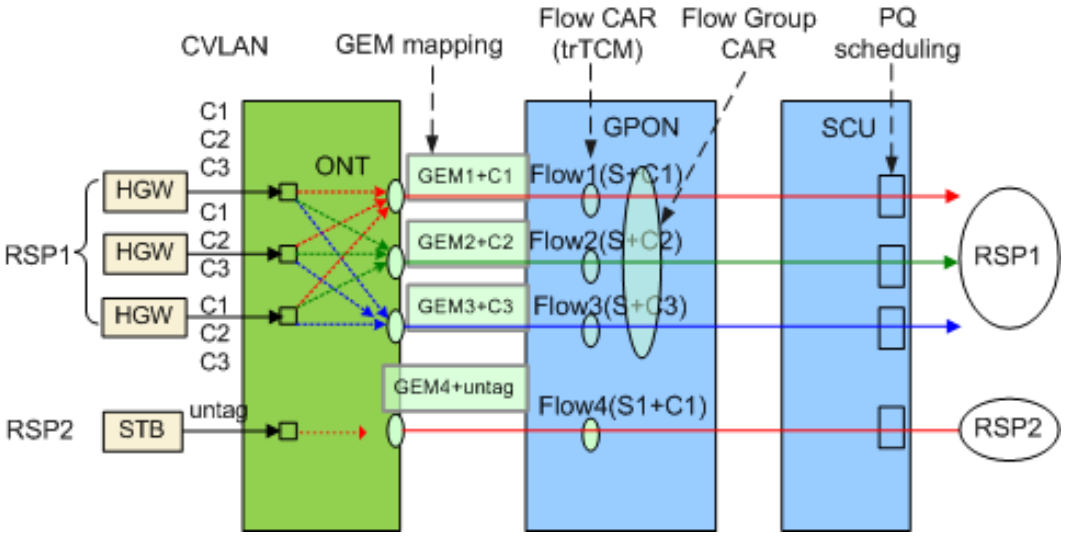


Implementation Solution

Figure 15-53 shows how HQoS is implemented based on a CAR group in the uplink direction.

In the downlink direction, a reverse process is performed. This topic only describes the implementation process in the uplink direction.

Figure 15-53 Flowchart for implementing HQoS based on a CAR group (in the uplink direction)



 Copyright © Huawei Technologies Co., Ltd.

RSP: retail service provider	HGW: home gateway
STB: set-top box	GEM: G-PON encapsulation method
trTCM: two rate three color marker	-

Table 15-14 Implementation principle of HQoS based on a CAR group in the uplink direction

Subject	Operation	Result
HGW	Identifies different types of services provided by the same RSP based on C-VLAN tags. In **Figure 15-53**, C1, C2, and C3 indicate different service types.	It receives various types of services and identifies them based on the C-VLAN tags of service packets.
ONT	Maps each type of service into one GEM port based on VLAN tags. In the figure, services tagged as C1 are mapped to GEM port 1 while services tagged as C2 are mapped to GEM port 2.	It identifies various service packets based on GEM port+C-VLAN tags.
	Binds each GEM port to one T-CONT. Performs priority queuing (PQ) on T-CONTs. (To prevent queuing differences, the class of service, or CoS, must have been uniformly configured.)	Guarantees the uplink bandwidth of each type of service on a T-CONT.

 Copyright © Huawei Technologies Co., Ltd.

Subject	Operation	Result
Service board of the OLT	Performs traffic classification and priority processing. Each GEM port+C-VLAN tag identifies one service flow. The priority of the service flow can be configured (if the user-side priority is not trusted), or directly duplicated from the user-side priority (if the user-side priority is trusted). Performs two-level CAR: 1. Flow CAR (rate limiting on each service flow) using the trTCM algorithm. For details, see **Traffic Policing Classification: Classification Based on Service Flows**. 2. Group CAR (rate limiting on all service flow groups provided by the same RSP) based on the packet color.	It identifies service flows through traffic classification, preparing for CAR. Flow CAR is based on the packet color that is marked based on the CIR and PIR. A packet is marked green (and will be transmitted) if the packet rate is lower than the CIR, is marked yellow (and will be transmitted) if the packet rate is higher than the CIR but is lower than or equal to the PIR, or is dropped if the packet rate is higher than the PIR. Group CAR is implemented using a single-token leaky bucket mechanism. With the single-token leaky bucket mechanism, drop thresholds can be set for different priorities of yellow packets to ensure that yellow packets with a higher priority can be preferentially transmitted. (The drop thresholds are the half of the CAR thresholds set using the **CAR-threshold** command). **NOTE** The PIR of a flow group must be higher than or equal to the sum of the CIRs of all flows. The two-level CAR mechanism guarantees the QoS for each type of service and for all services of a home user.
Uplink port of the OLT	Prevents congestion using a color-based early drop mechanism. Schedules queues (PQ scheduling) based on priorities.	The color-based early drop mechanism ensures that all green packets of a priority are transmitted. PQ scheduling ensures that packets with a higher priority preferentially occupy bandwidth resources.

NOTE

One traffic stream can belong to only one group; one group can contain a maximum of eight service flows; all service flows of a group should belong to the same xPON port.

 Copyright © Huawei Technologies Co., Ltd.

15.13.8.3 HQoS Implementation Principle on an xPON Board

HQoS on an xPON board supports the following scenarios:

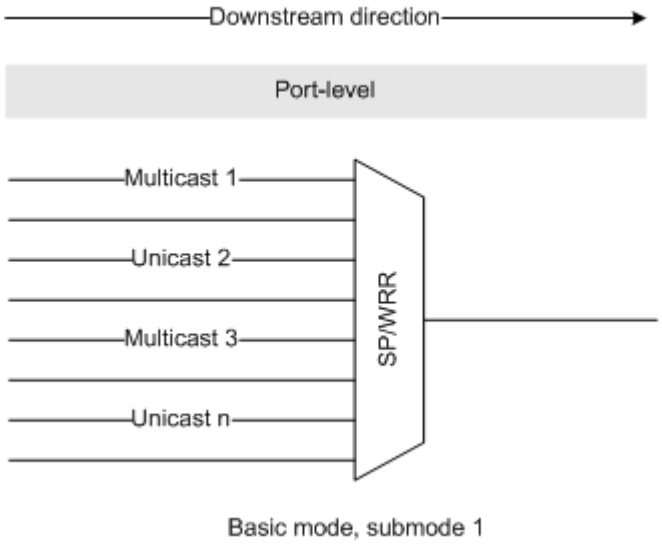
- Basic application
- Bandwidth priority
- Service priority

The following describes HQoS implementation principles in these 3 scenarios.

Basic Application Scenario

The basic mode inherits QoS functions on an xPON board. This mode has 2 sub-modes. **Figure 15-54** shows the HQoS implementation principle in sub-mode 1 of the basic mode.

Figure 15-54 Sub-mode 1 of basic mode



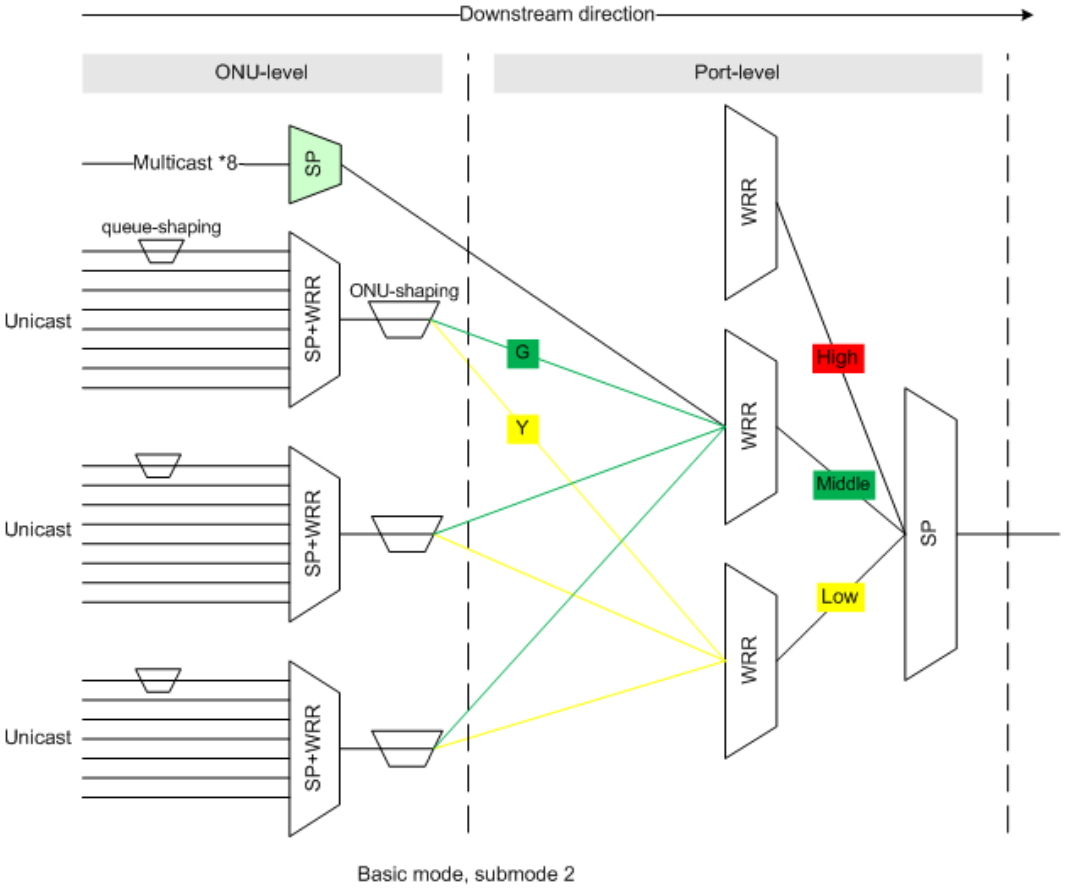
As shown in the figure, HQoS in this mode features the following:

- User unicast and multicast packets are enqueued on the PON port and are scheduled by SP and WRR by queue priorities.

When an xPON port has onu-shaping or queue-shaping, the HQoS mode is automatically switched to sub-mode 2 of the basic mode, as shown in **Figure 15-55**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-55 Sub-mode 2 of basic mode



HQoS in this mode features the following:

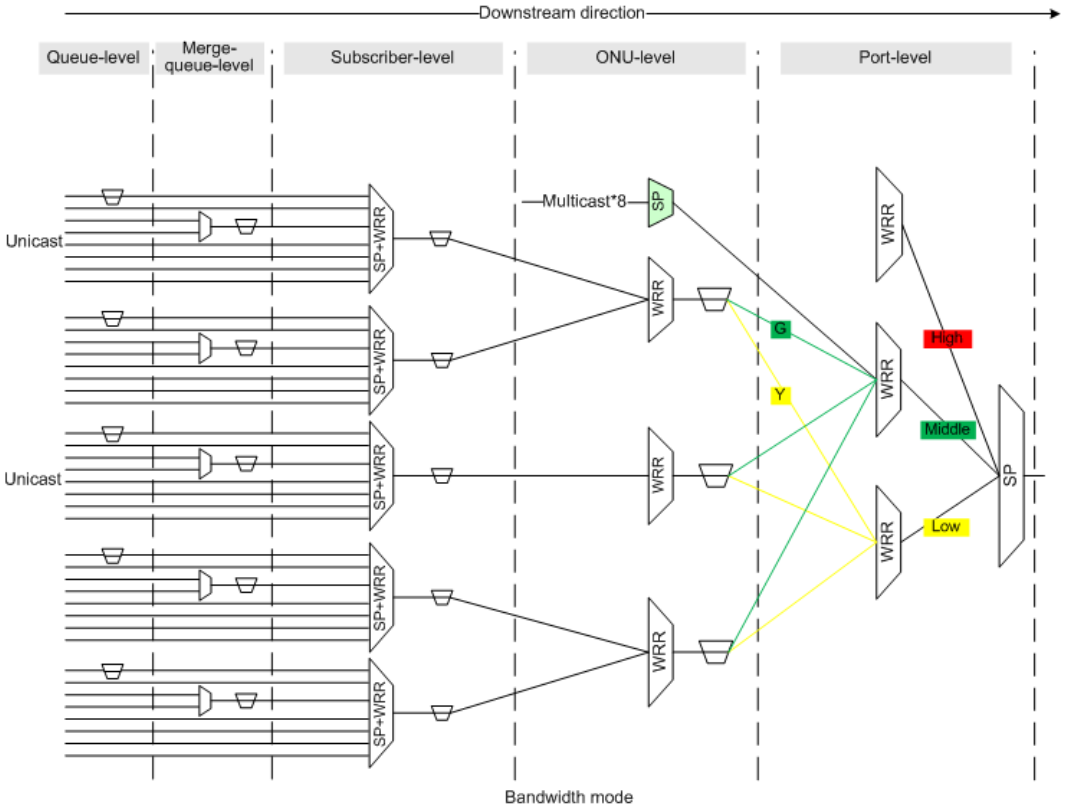
- Multicast packets are enqueued on multicast queues.
- Unicast packets are enqueued on 8 priority queues of the ONU.
- Multicast duplication packets and unicast packets are scheduled by SP.
- 8 priority queues inside an ONU are scheduled by SP and WRR.
- Every ONU supports 8 queue shapers with only valid PIR.
- Every ONU supports a shaper with valid CIR and PIR.

Bandwidth Priority Scenario

In the bandwidth priority mode, the system preferentially assures bandwidths at all hierarchies for fair user usage, as shown in **Figure 15-56**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-56 Bandwidth priority mode



HQoS in this mode features the following:

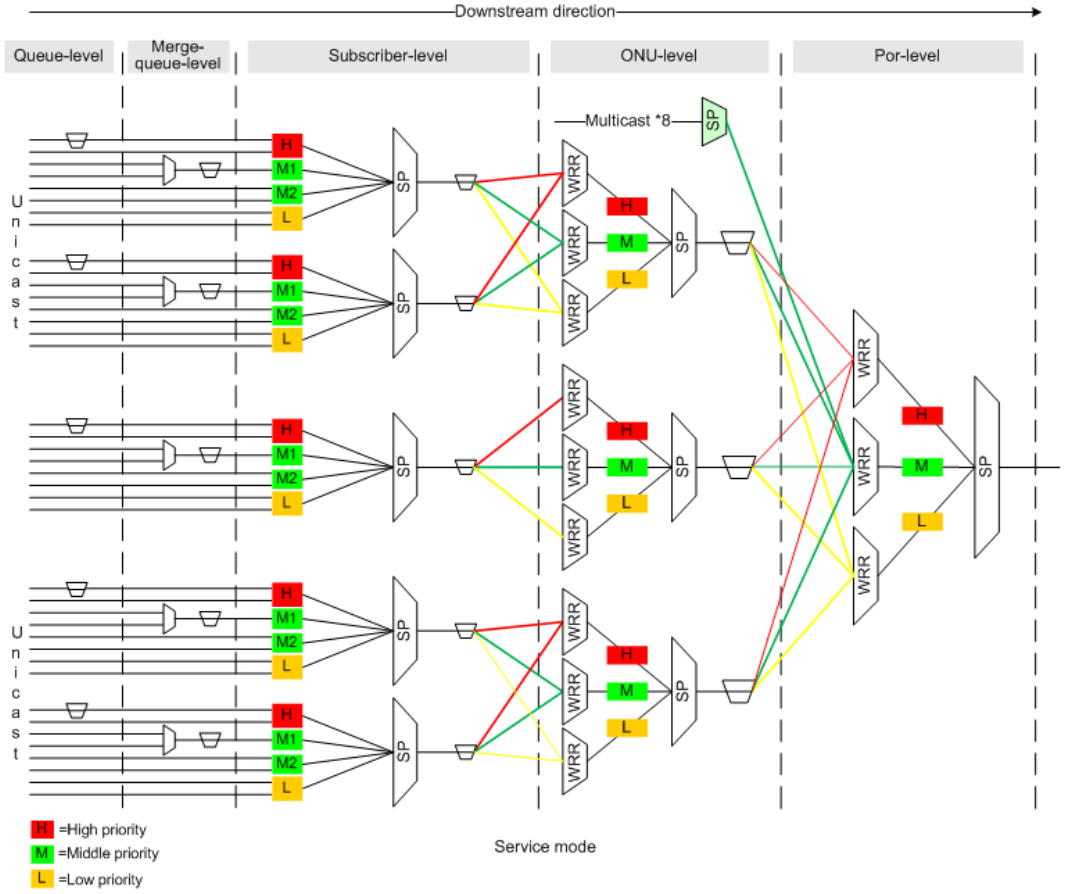
- Multicast packets are enqueued on multicast queues.
- Unicast packets are enqueued on 8 priority queues.
- (Multicast duplication packets, ONU CIR, and) yellow ONU packets are scheduled by SP on a PON port.
- Every ONU supports a shaper with valid CIR and PIR. Between ONUs, CIRs are scheduled by RR and EIRs (PIR-CIR) are scheduled by WRR (weights are calculated by EIR or obtained from the traffic profile).
- Users under an ONU are scheduled by WRR. Between users, CIRs are scheduled by RR and EIRs (PIR-CIR) are scheduled by WRR (weights are calculated by EIR or obtained from the traffic profile).
- Every user supports a shaper with valid CIR and PIR.
- Every user supports a merged queue shaper with only valid PIR.
- 8 priority queues inside users are scheduled by SP and WRR.
- 8 priority queues for users support queue shaper with only valid PIR.

Service Priority Scenario

In the service priority scenario, the system preferentially assures services with higher priorities for service scheduling, as shown in **Figure 15-57**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-57 Service priority mode



HQoS in this mode features the following:

- Multicast packets are enqueued on multicast queues.
- Unicast packets are enqueued on 8 priority queues.
- Packets with high priority, (multicast duplication packets and packets with medium priority, and) packets with low priority are scheduled by SP on a PON port.
- Every ONU supports a shaper with valid CIR and PIR.
- Packets having the same priority of different ONUs are scheduled by WRR.
- Packets having the same priority of different users under ONUs are scheduled by WRR (weights are calculated by PIR). Priorities high, medium, and low are supported for priority groups.
- Every user supports a shaper with only valid PIR.
- Every user supports a merged queue shaper with only valid PIR.
- 8 priority queues for users are scheduled by SP.
- 8 priority queues for users support queue shaper with only valid PIR.

15.13.8.4 HQoS Implementation Principle on a P2P Board

HQoS on a P2P board supports the following scenarios:

 Copyright © Huawei Technologies Co., Ltd.

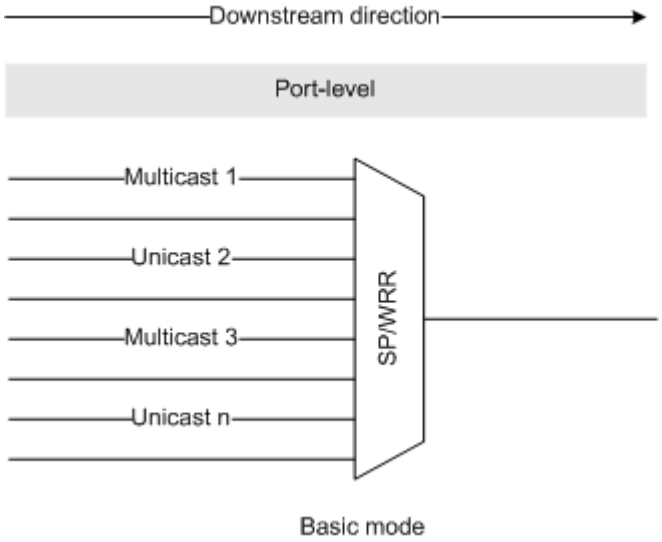
- Basic application
- Bandwidth priority
- Service priority

The following describes HQoS implementation principles in these 3 scenarios.

Basic Application Scenario

The basic mode inherits QoS functions on a P2P board. **Figure 15-58** shows the HQoS implementation principle in the basic mode.

Figure 15-58 Basic mode



As shown in the figure, HQoS in this mode features the following:

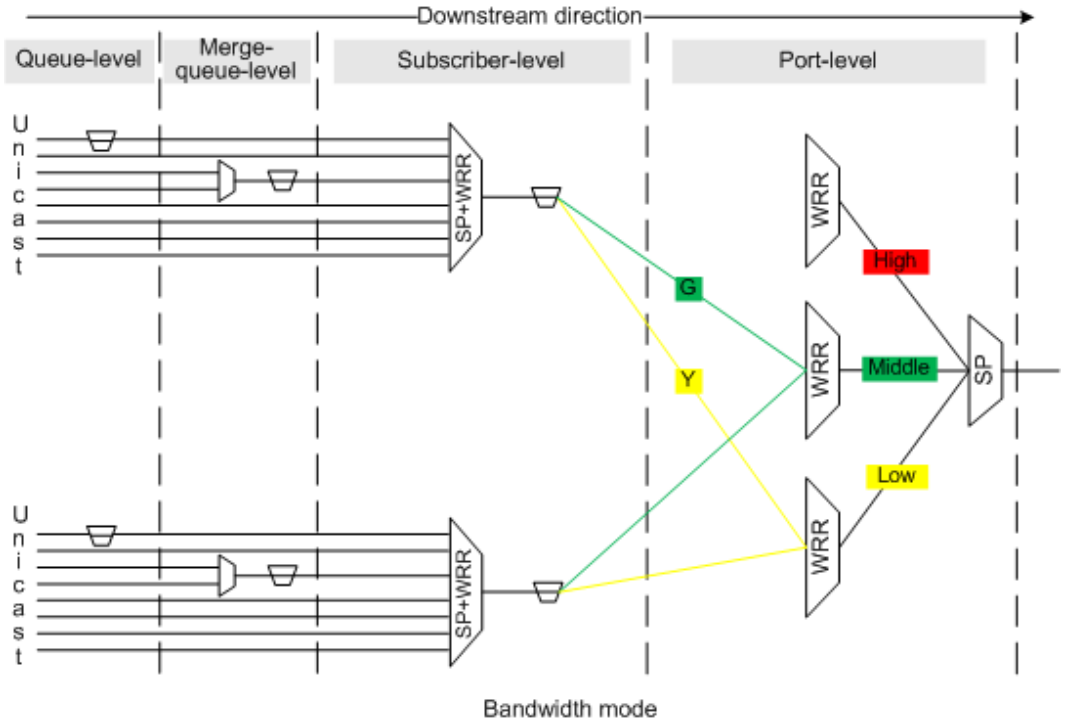
- User unicast and multicast packets are enqueued on the Ethernet port and are scheduled by SP and WRR by queue priorities.

Bandwidth Priority Scenario

In the bandwidth priority mode, the system preferentially assures bandwidths at all hierarchies for fair usage between users, as shown in **Figure 15-59**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-59 Bandwidth priority mode



HQoS in this mode features the following:

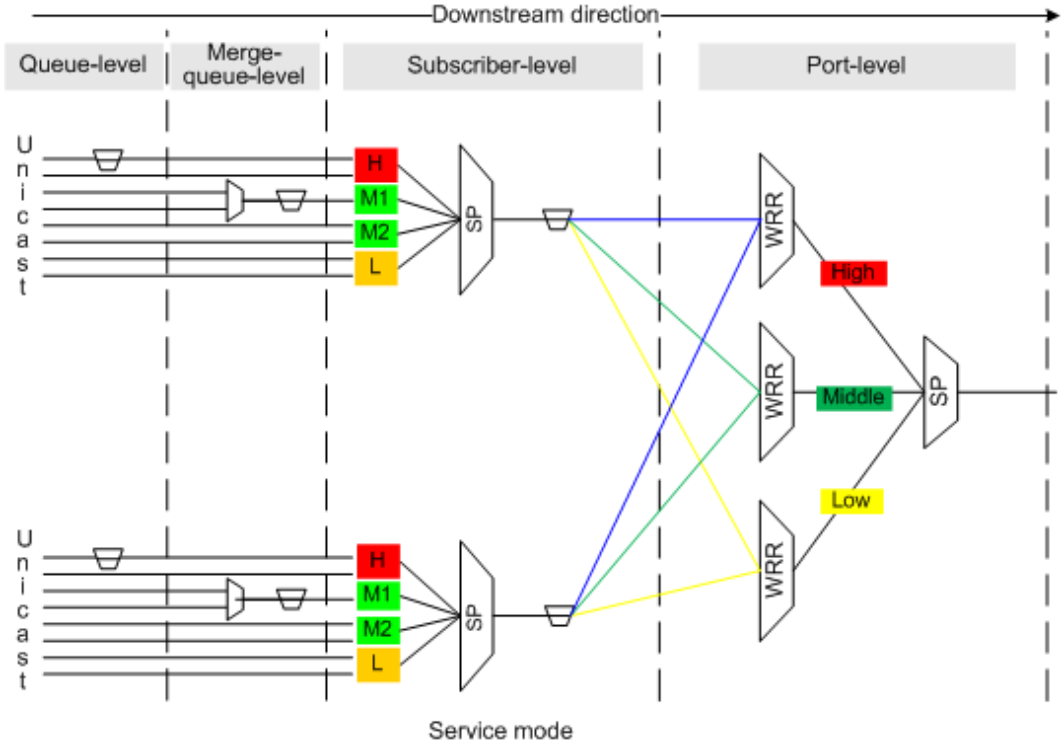
- Unicast packets are enqueued on 8 priority queues.
- Green and yellow packets are scheduled by SP on an Ethernet port.
- Every user supports a shaper with valid CIR and PIR. Green packets with CIR are scheduled between users by RR and yellow packets with EIR (PIR-CIR) are scheduled between users by WRR (weights are calculated by EIR).
- Every user supports a merged queue shaper with only valid PIR.
- 8 priority queues inside users are scheduled by SP and WRR.
- 8 priority queues for users support queue shaper with only valid PIR.

Service Priority Scenario

In the service priority scenario, the system preferentially assures services with higher priorities for service scheduling, as shown in **Figure 15-60**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-60 Service priority mode



HQoS in this mode features the following:

- Unicast packets are enqueued on 8 priority queues.
- Packets with high, medium, and low priorities are scheduled by SP on an Ethernet port.
- Every user supports a shaper with only valid PIR.
- Every user supports a merged queue shaper with only valid PIR.
- 8 priority queues for users are scheduled by SP.
- 8 priority queues for users support queue shaper with only valid PIR.

15.13.9 Networking Application of HQoS

Networking Application of HQoS based on Port+VLAN

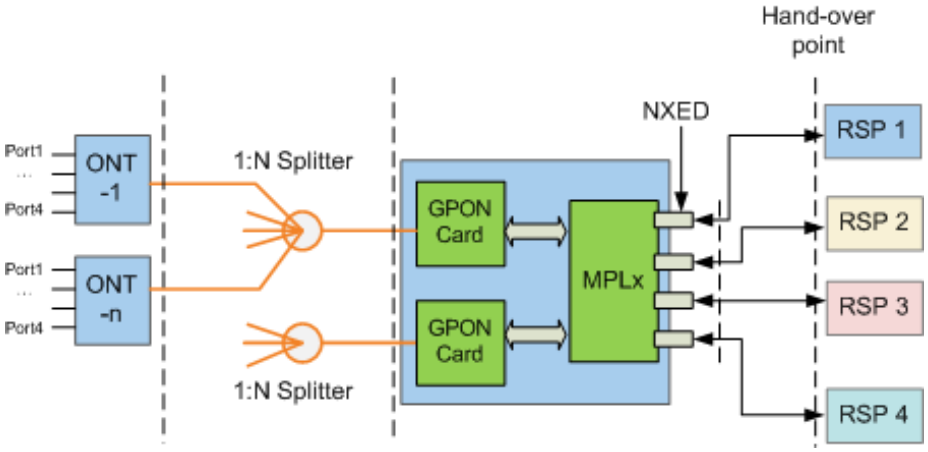
Figure 15-61 shows typical networking of hierarchical quality of service (HQoS) based on Port+VLAN. When Open Access is deployed, the carrier is only responsible for the network between the optical line terminal (OLT) and optical network terminal or ONT (including the ODN), while retail service providers (RSPs) are responsible for devices connected to the OLT and home area devices (an ONT provides 4 Ethernet UNI interfaces and each interface can be connected to one home device). The carrier network and RSP network are therefore interconnected at the uplink interface of the OLT.

 Copyright © Huawei Technologies Co., Ltd.

Generally, RSPs use different physical uplink interfaces. Sometimes, one RSP may use multiple physical uplink interfaces. If required, link convergence can be used to converge multiple uplink ports into one logical channel.

Since RSPs use different physical interfaces, each RSP can use all VLAN IDs, and duplicated VLAN IDs can be used for different RSPs.

Figure 15-61 Networking application of HQoS based on Port+VLAN



With the preceding networking, services that a carrier provides for RSPs cover the following types and the services address the following requirements:

- Voice services with symmetric uplink and downlink bandwidth, with CIR, but without PIR (voice services have the highest priority)
- Data services with symmetric uplink and downlink bandwidth, with CIR, but without PIR
- Data services with asymmetric uplink and downlink bandwidth (usually downlink bandwidth is far higher than the uplink bandwidth), with downlink PIR but without uplink PIR
- An ONT user port can carry traffic with multiple priorities:
  - Packets with a higher priority can preferentially occupy CIR resources. If there are remaining CIR resources, packets with a lower priority can be processed.
  - If the packet rate exceeds the CIR, packets with a higher priority can preferentially occupy PIR resources.
  - If there are remaining PIR resources, packets with a lower priority can be processed.
- For each type of data service, the CIR must be guaranteed. If required, PIR resources can be used in a certain proportion according to the network congestion status.

The preceding requirements can be well addressed by HQoS based on Port+VLAN.

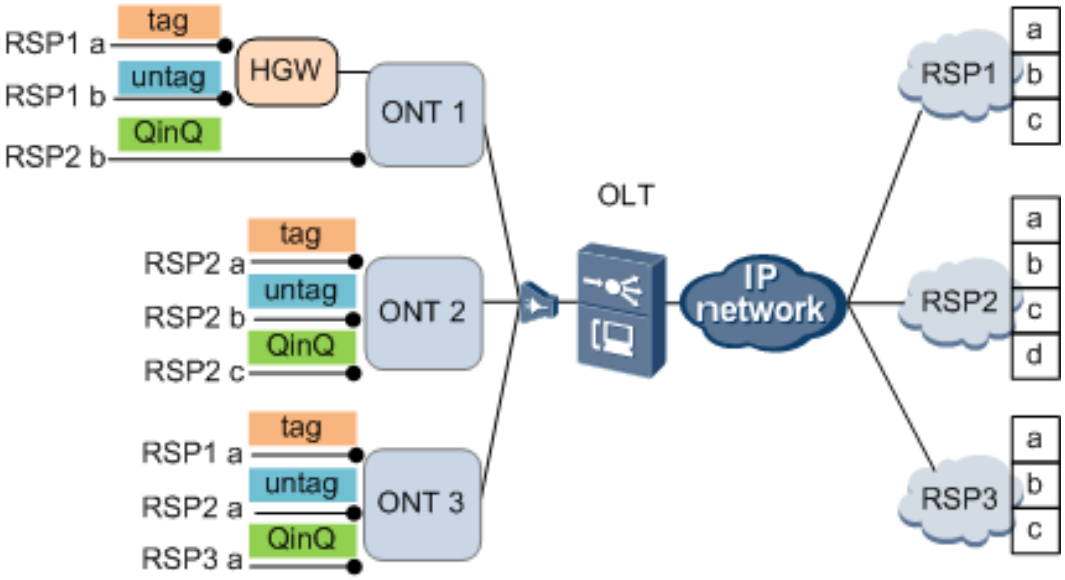
Networking Application of CAR-Group-based HQoS

As shown in **Figure 15-62**, users under the same ONT may belong to different RSPs. Different services of the same user are mapped to different service flows on

 Copyright © Huawei Technologies Co., Ltd.

the OLT. Given that different service CIRs/PIRs are guaranteed, the total bandwidth of each RSP needs to be ensured and each service should be allowed to occupy the total bandwidth when a burst occurs in the traffic. To put it simply, rate limitation needs to be performed on the RSP. To do so, a group based on all service flows of an RSP can be created, and then the total bandwidth of a user can be limited by limiting the bandwidth of the group. Such is a typical application of CAR group.

Figure 15-62 Networking application of HQoS based on a CAR group



Networking Application of xPON-board-based HQoS

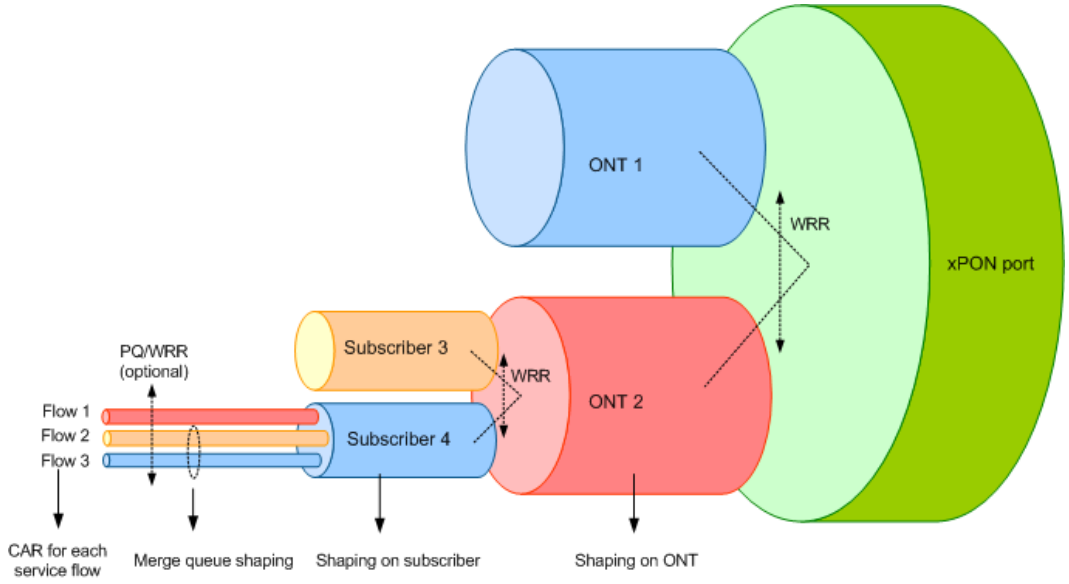
Figure 15-63 shows an HQoS application in the bandwidth priority mode of an xPON port. One xPON port connects to multiple ONTs, every ONT connects to multiple users, and every user is provisioned with multiple services. In this scenario, the bandwidth priority mode is used to ensure bandwidths of services at different hierarchies.

- CIR and EIR (PIR-CIR) permitted on different ONTs are allocated by preset weights.
- CIR and EIR (PIR-CIR) permitted by different users on an ONT are allocated by preset weights.
- WRR scheduling is used between users and between ONTs.

This application preferentially ensures user bandwidths for fair usage between users. It is applicable to the scenario in which the user priority is not a sensitive factor. In this case, when a network congestion occurs, the bandwidth plays an important role and users who subscribe to higher bandwidths have the high scheduling weight.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-63 Networking Application of xPON-board-based HQoS



Networking application of P2P-board-based HQoS

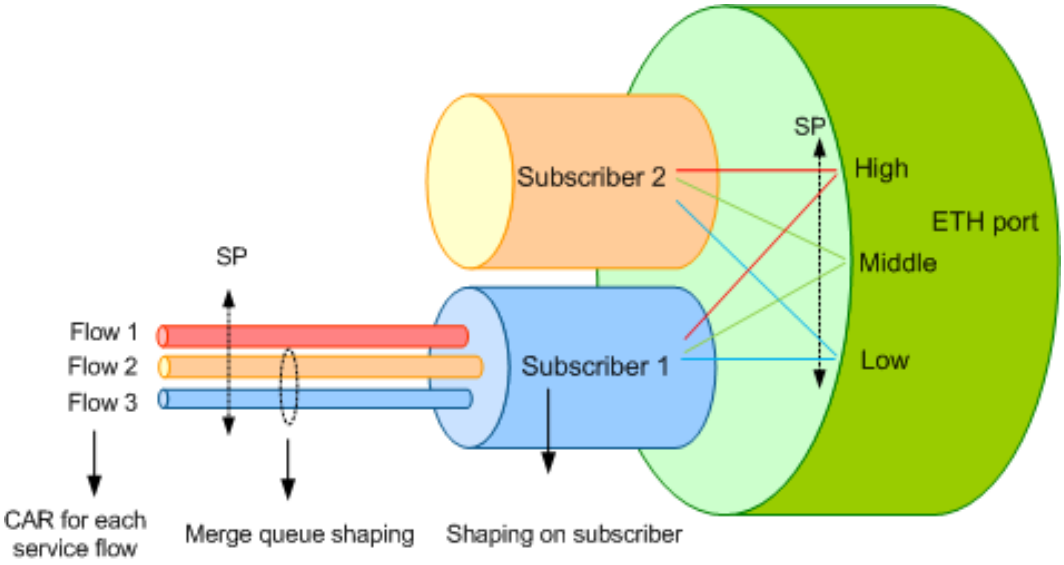
Figure 15-64 shows an HQoS application in the service priority mode of an Ethernet port. An Ethernet port connects to multiple users and every user is provisioned with multiple services. In this scenario, the service priority mode is used to ensure services with higher priority for service scheduling.

- 8 priority queues inside users are scheduled by SP.
- Packets with different priorities between users are scheduled by SP.

This application preferentially ensures services with higher priority for service scheduling. It is applicable to the scenario in which the user priority is a sensitive factor. In this case, when a network congestion occurs, the service priority plays an important role and services with higher priority are scheduled first.

 Copyright © Huawei Technologies Co., Ltd.

Figure 15-64 Networking application of P2P-board-based HQoS



15.13.10 Reference Standards and Protocols of HQoS

The following lists the reference standards and protocols of this feature:

RFC2698: A Two Rate Three Color Marker

15.13.11 Configuring HQoS

The hierarchical QoS (HQoS) is a QoS technology that controls user traffic on a port with finer granularity and also schedules services of a user based on the service priority. This topic describes the configuration of HQoS.

15.13.11.1 Configuring HQoS Based on Port+VLAN

This topic describes how to configure the Port+VLAN-based HQoS to ensure different CIRs and PIRs for data with different priorities in the private line services (that is, data with a higher priority preferentially occupies the bandwidth).

Prerequisites

Networking Application of HQoS based on Port+VLAN is configured.

Context

Configuration method:

1. Configure a traffic profile.
2. Use the traffic profile to limit the rate of service flows for HQoS users.
3. Limit the rate of HQoS users based on the CAR threshold and mark colors of packets having different priorities.

 Copyright © Huawei Technologies Co., Ltd.

4. Limit rates of HQoS user groups (port+SVLAN) based on colors.

Precaution:

HQoS is implemented through user configurations. Rate limitation on HQoS user groups is implemented using the single leaky bucket algorithm and bound TrTCM traffic profile. In the configured traffic profile, CIR = PIR; otherwise, PIR prevails for rate limitation. Ensure that PIR ≥ Total CIR bandwidths of all HQoS users.

Procedure

Step 1 Configure a traffic profile.

Run the **traffic table ip** command to configure the CIR and PIR of a service, and priority copy policy and enqueuing policy of packets.

Step 2 Configure rate limitation on Layer 2 service flows (HQoS).

Run the **service-port** command to configure an HQoS user and use the traffic profile to limit the rate of this user.

Step 3 Limit the rate of the HQoS user based on the CAR threshold and mark packet colors.

Run the **car-threshold profile** command to configure drop thresholds for packets having different priorities and mark yellow and green for packets.

Step 4 Limit the rate of an HQoS user group based on port+VLAN.

Run the **car-port portid vlan** command to specify a traffic profile to limit the rate of an HQoS user group.

NOTE

An HQoS user group can be considered as a collection of users on a specified port within a specified VLAN range.

- **inbound** *ip-traffic-table-index*: Sets the traffic profile index for the packet from the outside of the device to the inside of the device.
- **outbound** *ip-traffic-table-index*: Sets the traffic profile index for the packet from the inside of the device to the outside of the device.

Step 5 Query the configuration results.

1. Run the **display traffic table ip** command to query the traffic profile configurations.
2. Run the **display service-port** command to query the HQoS user configurations.
3. Run the **display car-threshold profile** command to query drop thresholds for packets having different priorities.
4. Run the **display car-port** command to query rate limitation configurations of an HQoS user group.

----End

 Copyright © Huawei Technologies Co., Ltd.

Example

In an open access network, a retail service provider (RSP) leases lines and bandwidths of a carrier and uses upstream port 0/2/0 on the upstream board to provide the following data service package for FTTH users:

- Assured bandwidth 4 Mbit/s, symmetric upstream and downstream bandwidths, and permitted burst bandwidth 6 Mbit/s.
- User-side VLAN 20, network-side VLAN 100, trust user-side priority, and queue scheduling in egress queues by packet priorities.
- Priorities 0, 3, and 6 for the data service, and packets having the higher priority preferentially use the assured bandwidth and burst bandwidth.
- Packets having the lower priorities are dropped first when a congestion occurs on a port. Drop thresholds for packets having priorities 0, 3, and 6 are 12%, 50%, and 100%.
- Total upstream bandwidth of the data service on the port is 50 Mbit/s and no burst bandwidth is permitted; total downstream bandwidth is 200 Mbit/s and the permitted burst bandwidth is 250 Mbit/s.

Configure HQoS in this scenario:

```
//Configures the traffic profile for the data service of residential users by the RSP.
huawei(config)#car-threshold profile 0 cos0 12 cos3 50 cos6 100
huawei(config)#traffic table ip index 8 cir 4096 pir 6144 priority user-cos 0 priority-policy tag-In-Package

//Configures the traffic profile for an RSP leased port by the carrier.
huawei(config)#traffic table ip index 9 cir 51200 car-threshold profile 0  priority user-cos priority-policy tag-In-Package  //Indicates the downstream traffic profile.
huawei(config)#traffic table ip index 10 cir 204800 pir 256000 car-threshold profile 0  priority user-cos priority-policy tag-In-Package   //Indicates the upstream traffic profile.

//Indicates an HQoS user (Layer 2 service flow).
huawei(config)#service-port 1 vlan 100 eth 0/2/0 multi-service user-vlan 20 rx-cttr 8 tx-cttr 8
huawei(config)#interface eth 0/2
huawei(config-if-eth-0/2)#car-port 0 vlan 100 inbound 10 outbound 9
```

After the configuration:

- Assured bandwidth 4 Mbit/s for the user when a network congestion occurs; burst bandwidth 6 Mbit/s for the user when the network is idle.
- Packets having different priorities share the bandwidth and packets are dropped from ones having the lower priorities when a network congestion occurs.
- Total bandwidth of the data service provided on the leased port by an RSP is at the most of the bandwidth allocated by the carrier.

15.13.11.2 Configuring HQoS Based on CAR Group

This topic describes how to configure HQoS based on CAR group for ensuring the bandwidth of each service of a user and the total bandwidth of the user.

Prerequisites

HQoS networking based CAR-Group is configured.

 Copyright © Huawei Technologies Co., Ltd.

Context

Configuration method:

1. Configure a traffic profile.
2. Use the traffic profile to limit the rate of various service flows for the same user.
3. All the user's service flows are added to a service flow group and rate limitation is performed

Precaution:

HQoS is implemented through user configurations. Rate limitation on CAR-group is implemented using the single leaky bucket algorithm and bound TrTCM traffic profile. In the configured traffic profile, CIR = PIR; otherwise, PIR prevails for rate limitation. Ensure that PIR in CAR-group ≥ Total CIR bandwidths of all service flows.

Procedure

Step 1 Configure a traffic profile.

Run the **traffic table ip** command to configure the CIR/PIR of a service/service group, priority copy policy, and enqueuing policy of packets.

Step 2 Configure rate limitation on service flows of a user.

Run the **service-port** command to configure different service flows for a user and use the traffic profile to limit the rate of the service flows.

Step 3 Configure CAR for an HQoS user group.

1. Create a CAR-group.

   Run the **car-group** command to create a CAR group for a service port and bind a traffic profile to the CAR group.

2. Add a service port to the CAR group.

   In the global config mode, run the **car-group add-member service-port** command to add a service port to the CAR group. Bandwidth of service ports in this CAR group is limited by that of the traffic profile bound to the CAR group.

   NOTE

   An HQoS user group can be considered as a collection of users whose port IDs and VLAN IDs are within the port+VLAN range specified by this command.

Step 4 Query the configuration results.

1. Run the **display traffic table ip** command to query the traffic profile configurations.
2. Run the **display service-port** command to query the configuration of the user's service flow.
3. Run the **display car-group** to query the configuration of the CAR group.

----End

 Copyright © Huawei Technologies Co., Ltd.

Example

In an open access network, a retail service provider (RSP) leases lines and bandwidths of a carrier to provide the following data service packages for FTTH users:

- Internet, voice, and video services are provided through the same GPON ONT.
- User-side VLANs of the Internet, voice, and video services are 10, 20, and 30 respectively. The network-side VLANs are 100, 200, and 300.
- Different services have different bandwidth and priority requirements.
  - The downstream assured bandwidth of the Internet service is 2 Mbit/s, the burst bandwidth is 4 Mbit/s, the upstream assured bandwidth is 512 Kbit/s, and no burst bandwidth exists. The priority is 0.
  - The assured upstream and downstream bandwidths of the voice service are 512 Kbit/s and no burst bandwidth exists. The priority is 7.
  - The assured upstream and downstream bandwidths of the video service are 4 Mbit/s and the burst bandwidth is 6 Mbit/s. The priority is 4.
- The total upstream and downstream bandwidths of the home users are limited to 6 Mbit/s and 8 Mbit/s respectively.

Configure HQoS in this scenario:

```
//Traffic profile
huawei(config)#traffic table ip index 8 cir 512 priority 0 priority-policy tag-In-Package //Upstream Internet service
huawei(config)#traffic table ip index 9 cir 2048 pir 4096 priority 0 priority-policy tag-In-Package //Downstream Internet service
huawei(config)#traffic table ip index 10 cir 512 priority 7 priority-policy tag-In-Package //Upstream and downstream voice service
huawei(config)#traffic table ip index 11 cir 4096 pir 6144 priority 4 priority-policy tag-In-Package //Upstream and downstream video service
huawei(config)#traffic table ip index 20 cir 6144 priority user-cos 0 priority-policy tag-In-Package //Total upstream bandwidth
huawei(config)#traffic table ip index 21 cir 8192 priority user-cos 0 priority-policy tag-In-Package //Total downstream bandwidth
//Internet service flow
huawei(config)#service-port 1 vlan 100 gpon 0/1/1 ont 1 gemport 1 multi-service user-vlan 10 rx-cttr 9 tx-cttr 8
//Voice service flow
huawei(config)#service-port 2 vlan 200 gpon 0/1/1 ont 1 gemport 2 multi-service user-vlan 20 rx-cttr 10 tx-cttr 10
//Video service flow
huawei(config)#service-port 3 vlan 300 gpon 0/1/1 ont 1 gemport 3 multi-service user-vlan 30 rx-cttr 11 tx-cttr 11
CAR for an HQoS user group
huawei(config)#car-group 1 inbound traffic-table index 20 outbound traffic-table index 21
huawei(config)#car-group 1 add-member service-port 1-3
```

After the configuration:

- The bandwidth of each service of the user can be guaranteed. When the network is idle, the Internet and video service can obtain certain burst bandwidth.
- The overall bandwidth of the user is guaranteed.

15.13.11.3 Configuring HQoS for an xPON Board

In an xPON access scenario, select a proper HQoS mode according to actual QoS requirements.

 Copyright © Huawei Technologies Co., Ltd.

Context

xPON boards support HQoS applications in the following exclusive scenarios:

- Basic application scenario: Inherits the original QoS function. It is the default mode.
- Bandwidth priority scenario: The system preferentially assures bandwidths at all hierarchies for fair user usage.
- Service priority scenario: The system preferentially assures services with higher priorities for service scheduling.
- Slice priority scenario: The system preferentially assures slices are scheduled according to their priorities.

For detailed application scenarios, see **15.13.8.3 HQoS Implementation Principle on an xPON Board**.

Procedure

- **Configure HQoS in a basic application scenario.**
  a. In xPON mode, run the **hqos mode *basic*** command to set the HQoS mode of an xPON board to basic.
  b. Run the **display hqos mode** command to query the HQoS mode of a board.
- **Configure HQoS in a bandwidth priority scenario.**
  a. Set the HQoS mode to **bandwidth**.

     In xPON mode, run the **hqos mode *bandwidth*** command to set the HQoS mode of an xPON board to bandwidth priority.
  b. Configure rate limitation on an ONT.

     Run the **traffic-limit ont** command to limit the traffic of downstream packets on a specified ONT.
  c. Configure the HQoS user profile and its attributes.
     i. Run the **subscriber profile** command to create an HQoS user profile.
     ii. Run the **subscriber shaping** command to limit the rate of an HQoS user.
  d. (Optional) Configure the merged queue profile of HQoS users and its attributes.
     i. Run the **merge queue profile** command to create a merged queue profile and bind queues.
     ii. Run the **merge queue bind** command to bind a merged queue profile.
     iii. Run the **merge queue shaping** command to set rate limitation on the merged queue.
  e. Add an HQoS user and limit the rate of this user.
     i. Run the **subscriber add** command to bind a user to the HQoS user profile and limit the rate of this user.
     ii. Run the **subscriber member add** command to add a service port as an HQoS user.

 Copyright © Huawei Technologies Co., Ltd.

f. Query the configuration.

- Run the **display hqos mode** command to query the HQoS mode of a board.
- Run the **display subscriber** command to query the HQoS user information.
- (Optional) Run the **display merge queue profile** command to query the merged queue information about an HQoS user.

- **Configure HQoS in a service priority scenario.**

  NOTE

  HQoS configuration in this scenario is the same as that in the bandwidth priority scenario. The only difference is that HQoS mode in this scenario is set to **service**.

  a. Set the HQoS mode to **service**.

     In xPON mode, run the **hqos mode *service*** command to set the HQoS mode of an xPON board to service priority.

  b. Configure rate limitation on an ONT.

     Run the **traffic-limit ont** command to limit the traffic of downstream packets on a specified ONT.

  c. Configure the HQoS user profile and its attributes.

     i. Run the **subscriber profile** command to create an HQoS user profile.
     ii. Run the **subscriber shaping** command to limit the rate of an HQoS user.

  d. (Optional) Configure the merged queue profile of HQoS users and its attributes.

     i. Run the **merge queue profile** command to create a merged queue profile and bind queues.
     ii. Run the **merge queue bind** command to bind a merged queue profile.
     iii. Run the **merge queue shaping** command to set rate limitation on the merged queue.

  e. Add an HQoS user and limit the rate of this user.

     i. Run the **subscriber add** command to bind a user to the HQoS user profile and limit the rate of this user.
     ii. Run the **subscriber member add** command to add a service port as an HQoS user.

  f. Query the configuration.

     - Run the **display hqos mode** command to query the HQoS mode of a board.
     - Run the **display subscriber** command to query the HQoS user information.
     - (Optional) Run the **display merge queue profile** command to query the merged queue information about an HQoS user.

----End

 Copyright © Huawei Technologies Co., Ltd.

Example

The following is an example of the configurations used to plan 3 ONTs under GPON port 0/1/0:

- Set the bandwidth priority mode.
- Configure a user for ONT 1 (ONT ID 1).
  - Configure 2 service flows: the rate of one service flow with index 100 and priority 4 is limited to 600 Mbit/s and that of the other one with index 101 and priority 0 is limited to 150 Mbit/s.
  - Limit the shaping rate of users to 800 Mbit/s (traffic profile index 10, CIR 400 Mbit/s, and PIR 800 Mbit/s).
  - Limit the shaping rate of ONT 1 to 1 Gbit/s (traffic profile index 11, CIR 400 Mbit/s, and PIR 1 Gbit/s).
- Configure ONT 2 (ONT ID 2) in the same way as ONT 1 (the indexes of the two service flows are 102 and 103).
- Configure 2 users for ONT 3 (ONT ID 3) under 2 GE ports on the ONT.
  - Configure 2 service flows for user 1 (ID 2): the rate of one service flow with index 102 and priority 4 is limited to 600 Mbit/s and that of the other one with index 103 and priority 0 is limited to 150 Mbit/s.
  - Limit the shaping rate of user 1 to 800 Mbit/s (traffic profile index 12, CIR 600 Mbit/s, and PIR 800 Mbit/s).
  - Configure 2 service flows for user 2 (ID 3): the rate of one service flow with index 104 and priority 4 is limited to 600 Mbit/s and that of the other one with index 105 and priority 0 is limited to 150 Mbit/s.
  - Limit the shaping rate of user 2 to 1 Gbit/s (CIR 400 Mbit/s and PIR 1 Gbit/s).
  - Limit the shaping rate of ONT 3 to 2 Gbit/s (traffic profile index 13, CIR 1.2 Gbit/s, and PIR 2 Gbit/s).

The configuration procedures are as follows:

```
 //Configure the HQoS mode.

huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#hqos mode bandwidth
huawei(config-if-gpon-0/1)#quit

..//Cofnigure the traffic profile.
huawei(config)#traffic table ip index 10 cir 409600 pir 819600 priority user-cos 0 priority-policy tag-In-package
huawei(config)#traffic table ip index 11 cir 409600 pir 1024000 priority user-cos 0 priority-policy tag-In-package
huawei(config)#traffic table ip index 12 cir 614400 pir 819600 priority user-cos 0 priority-policy tag-In-package
huawei(config)#traffic table ip index 13 cir 1228800 pir 2048000 priority user-cos 0 priority-policy tag-In-package

 //Configure the profile for HQoS users under ONT 1 and ONT 2.
huawei(config)#subscriber profile profile-id 1
huawei(config-subscriber-profile-1)#subscriber shaping outbound traffic-table index 10
huawei(config-subscriber-profile-1)#quit

 //Add HQoS users under ONT 1 and ONT 2.
huawei(config)#subscriber add 0/1/0 ont 1 1 profile-id 1
huawei(config)#subscriber member add 0/1/0 ont 1 1 service-port 100
huawei(config)#subscriber member add 0/1/0 ont 1 1 service-port 101
huawei(config)#subscriber add 0/1/0 ont 2 1 profile-id 1
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#subscriber member add 0/1/0 ont 2 1 service-port 102
huawei(config)#subscriber member add 0/1/0 ont 2 1 service-port 103

  //Limit rates of ONT 1 and ONT 2.
huawei(config)#traffic-limit ont 0/1/0 1 down-stream traffic-table index 11
huawei(config)#traffic-limit ont 0/1/0 2 down-stream traffic-table index 11

 //Configure the profile for HQoS user 1 under ONT 3.
huawei(config)#subscriber profile profile-id 2
huawei(config-subscriber-profile-2)#subscriber shaping outbound traffic-table index 12
huawei(config-subscriber-profile-2)#quit

  //Add HQoS user 1 under ONT 3.
huawei(config)#subscriber add 0/1/0 ont 3 2 profile-id 2
huawei(config)#subscriber member add 0/1/0 ont 2 2 service-port 102
huawei(config)#subscriber member add 0/1/0 ont 2 2 service-port 103

  //Configure the profile for HQoS 2 under ONT 3.
huawei(config)#subscriber profile profile-id 3
huawei(config-subscriber-profile-3)#subscriber shaping outbound traffic-table index 11
huawei(config-subscriber-profile-3)#quit

  //Add HQoS user 3 under ONT 3.
huawei(config)#subscriber add 0/1/0 ont 3 3 profile-id 3
huawei(config)#subscriber member add 0/1/0 ont 3 3 service-port 104
huawei(config)#subscriber member add 0/1/0 ont 3 3 service-port 105

  //Limit the rate of ONT 3.
huawei(config)#traffic-limit ont 0/1/0 3 down-stream traffic-table index 13
```

Follow-up Procedure

Expected results:

- The CIR ratio of ONT 1, ONT 2, and ONT 3 is 1:1:3, and the total CIR of these 3 ONTs is 2.0 Gbit/s, which does not exceed the downstream bandwidth of a PON port.
- The EIR (PIR-CIR) ratio of ONT 1, ONT 2, and ONT 3 is 3:3:4, and the remaining bandwidth of a PON port can be used by EIR.
- ONT 1 permits bandwidth about 550 Mbit/s, containing CIR 400 Mbit/s for packets with priority 4 and EIR 150 Mbit/s for packets with priority 4.
- ONT 2 permits bandwidth about 550 Mbit/s, containing CIR 400 Mbit/s for packets with priority 4 and EIR 150 Mbit/s for packets with priority 4.
- ONT 3 permits bandwidth about 1400 Mbit/s:
  - The CIR ratio of user 1 and user 2 is 6:4 and EIR ratio is 1:3.
  - Packet bandwidth of user 1 is 700 Mbit/s, containing CIR 600 Mbit/s for packets with priority 4 and EIR 100 Mbit/s for packets with priority 0.
  - Packet bandwidth of user 2 is 700 Mbit/s, containing CIR 400 Mbit/s for packets with priority 4 and EIR 300 Mbit/s for packets with priority 4 (200 Mbit/s) and with priority 0 (100 Mbit/s).

15.13.11.4 Configuring HQoS for a P2P Board

In a P2P access scenario, select a proper HQoS mode according to actual QoS requirements.

 Copyright © Huawei Technologies Co., Ltd.

Context

P2P boards support HQoS applications in the following exclusive scenarios:

- Basic application scenario: Inherits the original QoS function. It is the default mode.
- Bandwidth priority scenario: The system preferentially assures bandwidths at all hierarchies for fair user usage.
- Service priority scenario: The system preferentially assures services with higher priorities for service scheduling.

For detailed application scenarios, see **15.13.8.4 HQoS Implementation Principle on a P2P Board**.

Procedure

- **Configure HQoS in a basic application scenario.**
  a. In ETH mode. run the **hqos mode *basic*** command to set the HQoS mode of an P2P board to basic.
  b. Run the **display hqos mode** command to query the HQoS mode of a board.
- **Configure HQoS in a bandwidth priority scenario.**
  a. Set the HQoS mode to **bandwidth**.

     In ETH mode. run the **hqos mode *bandwidth*** command to set the HQoS mode of an P2P board to bandwidth priority.
  b. Configure rate limitation for an Ethernet port.

     The **line-rate** command is used to limit the traffic rate of a specified Ethernet port.
  c. Configure the HQoS user profile and its attributes.
     i. Run the **subscriber profile** command to create an HQoS user profile.
     ii. Run the **subscriber shaping** command to limit the rate of an HQoS user.
  d. (Optional) Configure the merged queue profile of HQoS users and its attributes.
     i. Run the **merge queue profile** command to create a merged queue profile and bind queues.
     ii. Run the **merge queue bind** command to bind a merged queue profile.
     iii. Run the **merge queue shaping** command to set rate limitation on the merged queue.
  e. Add an HQoS user and limit the rate of this user.
     i. Run the **subscriber add** command to bind a user to the HQoS user profile and limit the rate of this user.
     ii. Run the **subscriber member add** command to add a service port as an HQoS user.
  f. Query the configuration.

 Copyright © Huawei Technologies Co., Ltd.

- Run the **display hqos mode** command to query the HQoS mode of a board.
- Run the **display subscriber** command to query the HQoS user information.
- (Optional) Run the **display merge queue profile** command to query the queue group information about an HQoS user.

- **Configure HQoS in a service priority scenario.**

  NOTE

  HQoS configuration in this scenario is the same as that in the bandwidth priority scenario. The only difference is that HQoS mode in this scenario is set to **service**.

  a. Set the HQoS mode to **service**.

     In ETH mode. run the **hqos mode *service*** command to set the HQoS mode of an P2P board to service priority.

  b. Configure rate limitation for an Ethernet port.

     The **line-rate** command is used to limit the traffic rate of a specified Ethernet port.

  c. Configure the HQoS user profile and its attributes.

     i. Run the **subscriber profile** command to create an HQoS user profile.

     ii. Run the **subscriber shaping** command to limit the rate of an HQoS user.

  d. (Optional) Configure the merged queue profile of HQoS users and its attributes.

     i. Run the **merge queue profile** command to create a merged queue profile and bind queues.

     ii. Run the **merge queue bind** command to bind a merged queue profile.

     iii. Run the **merge queue shaping** command to set rate limitation on the merged queue.

  e. Add an HQoS user and limit the rate of this user.

     i. Run the **subscriber add** command to bind a user to the HQoS user profile and limit the rate of this user.

     ii. Run the **subscriber member add** command to add a service port as an HQoS user.

  f. Query the configuration.

     - Run the **display hqos mode** command to query the HQoS mode of a board.
     - Run the **display subscriber** command to query the HQoS user information.
     - (Optional) Run the **display merge queue profile** command to query the queue group information about an HQoS user.

----End

 Copyright © Huawei Technologies Co., Ltd.

Example

The following is an example of the configurations used to plan 2 users under port 0/2/0 on the P2P board:

- Set the service priority mode.
- Limit the rate in the outbound direction of the port to 800 Mbit/s.
- Configure 3 service flows for user 1 (ID 1): the rate of one service flow with index 100 and priority 5 is limited to 300 Mbit/s, one with index 101 and priority 4 is limited to 250 Mbit/s, and one with index 102 and priority 0 is limited to 100 Mbit/s.
- Limit the shaping rate of user 1 to 600 Mbit/s (traffic profile index 10, CIR 400 Mbit/s, and PIR 600 Mbit/s).
- Limit the merged rate on queues 4 and 5 of user 1 to 500 Mbit/s (traffic profile index 11, CIR 300 Mbit/s, and PIR 500 Mbit/s).
- Configure 2 service flows for user 2 (ID 2): the rate of one service flow with index 103 and priority 5 is limited to 300 Mbit/s and that of the other one with index 104 and priority 4 is limited to 250 Mbit/s.
- Limit the shaping rate of user 2 to 400 Mbit/s (traffic profile index 12, CIR 300 Mbit/s, and PIR 400 Mbit/s).

The configuration procedures are as follows:

```
  //Configure the HQoS mode.
huawei(config)#interface gpon 0/2
huawei(config-if-eth-0/2)#hqos mode service
  This operation will reset the board, which may take several minutes. Are you
sure to continue ? (y/n)[n]:y
huawei(config-if-eth-0/2)#quit

..//Cofnigure the traffic profile.
huawei(config)#traffic table ip index 10 cir 409600 pir 614400 priority user-cos 0 priority-policy tag-In-package
huawei(config)#traffic table ip index 11 cir 307200 pir 512000 priority user-cos 0 priority-policy tag-In-package
huawei(config)#traffic table ip index 12 cir 307200 pir 409600 priority user-cos 0 priority-policy tag-In-package


  //Configure the profile for HQoS user 1.
huawei(config)#subscriber profile profile-id 1
huawei(config-subscriber-profile-1)#subscriber shaping outbound traffic-table index 10
huawei(config-subscriber-profile-1)#merge queue bind profile 3  //Binds the default profile 3 and uses queues 4 and 5.
huawei(config-subscriber-profile-1)#merge queue shaping outbound traffic-table index 11
huawei(config-subscriber-profile-1)#quit

  //Configure the profile for HQoS user 2.
huawei(config)#subscriber profile profile-id 2
huawei(config-subscriber-profile-2)#subscriber shaping outbound traffic-table index 12
huawei(config-subscriber-profile-2)#quit

  //Add HQoS user 1.
huawei(config)#subscriber add 0/2/0 1 profile-id 1
huawei(config)#subscriber member add 0/2/0 1 service-port 100
huawei(config)#subscriber member add 0/2/0 1 service-port 101
huawei(config)#subscriber member add 0/2/0 1 service-port 102

  //Add HQoS user 2.
huawei(config)#subscriber add 0/2/0 2 profile-id 2
huawei(config)#subscriber member add 0/2/0 2 service-port 103
huawei(config)#subscriber member add 0/2/0 2 service-port 104
```

 Copyright © Huawei Technologies Co., Ltd.

```
  //Limit the rate of Ethernet port 0/2/0.
huawei(config)#line-rate outbound 819200
```

15.14 End-to-End QoS

15.14.1 FTTH End-to-End QoS Policy

The FTTH end-to-end (E2E) QoS solution uses the differentiated service (DiffServ) model. In this solution, QoS parameters are carried in each packet header to ensure QoS guarantees. The following describes upstream and downstream FTTH E2E QoS applications.

Upstream E2E QoS

Upstream E2E QoS is shown in the **Figure 15-65** and **Table 15-15**.

1. QoS policy on the ONT.
   a. An ONT classifies traffic based on user ports or user-side VLAN IDs to distinguish services.
   b. Re-marks the 802.1p priority for service packets.
   c. Based on packet priorities, the ONT arranges packets into different priority queues by using PQ scheduling. This ensures that services with higher QoS requirements are forwarded preferentially.
   d. Limits the upstream traffic on user-side ports. (optional).
2. QoS policy on the OLT.
   a. Classifies traffic based on VLAN IDs and the 802.1p priority
   b. Trusts the user-side priority or re-marks the 802.1p priority for service packets.
   c. Implements priority scheduling and congestion management based on the 802.1p priority.
   d. Implements DBA to limit the ONUs upstream bandwidth.
   e. (Optional) limits upstream traffic bandwidth based on service flows.
3. The layer 2 metropolitan area network (MAN) between the OLT and the BRAS/service router (SR) implements priority scheduling based on the 802.1p priority of service packets.
4. Traffic classification is enabled on the ingress of the BRAS/SR to identify 802.1p values of different packets and re-mark packets according to the upper network.
   – When the upper-layer network is a native IP Layer 3 network, the BRAS/SR maps the 802.1p priority to the ToS/DSCP value in the IP header.
   – When the upper-layer network is an MPLS VPN, the BRAS/SR maps the 802.1p priority to MPLS EXP.

   (Optional) The BRAS/SR can also limit bandwidth for upstream user traffic based on service flows.

 Copyright © Huawei Technologies Co., Ltd.

5. Upper-level networks of the BRAS/SR implement traffic classification, congestion management, and congestion avoidance based on ToS/DSCP in the IP header or the MPLS EXP priority.

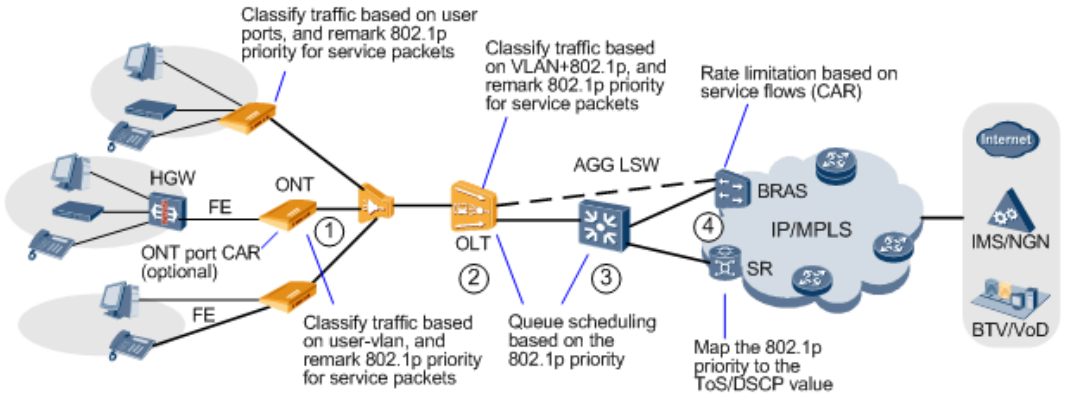
Table 15-15 Upstream E2E QoS policy

Device	QoS policy			
	Traffic classification	**Priority processing**	**Traffic policing**	**Queue scheduling**
ONT	• Based on user ports • Based on user-side VLAN IDs	Re-marks the 802.1p priority	(Optional) Limits the upstream traffic on user-side ports.	PQ schedule based on 802.1p priority
OLT	VLAN +802.1p priority	• Trusts the user-side priority • Re-marks the 802.1p priority	Implements DBA to limit the ONUs upstream bandwidth. (Optional) Limits upstream traffic bandwidth based on service flows.	PQ schedule based on 802.1p priority
L2 MAN	N/A	N/A	N/A	PQ schedule based on 802.1p priority of service packets

 Copyright © Huawei Technologies Co., Ltd.

Device	QoS policy			
	Traffic classification	Priority processing	Traffic policing	Queue scheduling
BRAS/SR	Based on VLAN	• Maps the 802.1p priority to the ToS/DSCP value in the IP header when upper-layer is IP L3 network. • Maps the 802.1p priority to MPLS EXP when upper-layer is MPLS VPN.	(Optional) Limits bandwidth for upstream user traffic based on service flows.	N/A
L3 network (IP/MPLS)	IP network: based on ToS/DSCP MPLS network: based on MPLS EXP	Trusts priority of packets	N/A	PQ

Figure 15-65 FTTH E2E QoS networking



 Copyright © Huawei Technologies Co., Ltd.

Downstream E2E QoS

Downstream E2E QoS is shown in the **Figure 15-65** and **Table 15-16**.

1. The packets of the IPTV, VoIP, and broadband Internet services are assigned different DSCP/MPLS EXP values based on priority requirements using related egress switches or provider edge (PE) devices. Core networks and backbone networks perform priority-based queue scheduling according to DSCP/MPLS EXP.
2. (Optional) Downstream packets are transmitted to edge BRAS/SR of the Layer 3 network through core networks and backbone networks. The BRAS/SR maps IP DSCP or MPLS EXP of packets to the VLAN 802.1p priority. The BRAS/SR can also limit bandwidth for downstream user traffic based on service flows.
   - For broadband Internet services in PPPoE mode, the BRAS controls the access bandwidth of each broadband user based on the rate authorized by the RADIUS server. For example, the rate can be 512 Kbit/s, 1 Mbit/s, or 2 Mbit/s.
   - Due to traffic jitter of the server or network, queue shaping is necessary for IPTV services on service routers.
3. Metropolitan networks and Layer 2 access networks downstream to the BRAS/SR implement priority scheduling based on the packet 802.1p priority.
4. (Optional) The OLT classifies traffic based on VLAN IDs and the 802.1p priority, and it trusts the network-side priority or re-marks the 802.1p priority for service packets. It implements priority scheduling and congestion management based on the 802.1p priority. It can limit downstream traffic bandwidth based on service flows.
5. (Optional) The ONT trusts the network-layer 802.1p priority and implements priority scheduling based on the 802.1p priority. The ONT can also limit downstream traffic bandwidth based on user-side ports.

 Copyright © Huawei Technologies Co., Ltd.

Table 15-16 Downstream E2E QoS policy

Device	QoS policy			
	Traffic classification	**Priority processing**	**Traffic policing**	**Queue scheduling**
L3 network (IP/MPLS)	N/A	Assign different priority values based on service types and networks. • IP network: assign different IP ToS/DSCP value • MPLS network: assign different MPLS EXP value	N/A	PQ
BRAS/SR	N/A	• Maps the ToS/DSCP value to 802.1p priority when upper-layer is IP L3 network. • Maps the MPLS EXP to 802.1p priority when upper-layer is MPLS network.	(Optional) Limits bandwidth for downstream user traffic based on service flows. • HSI service in PPPoE mode: CAR • IPTV service: queue shaping	N/A
L2 MAN	N/A	N/A	N/A	PQ schedule based on 802.1p priority of service packets
OLT	VLAN +802.1p priority	• Trusts the network-layer 802.1p priority • Re-marks 802.1p priority	(Optional) Limits downstream traffic bandwidth based on service flows.	PQ schedule based on 802.1p priority

 Copyright © Huawei Technologies Co., Ltd.

Device	QoS policy			
	Traffic classification	Priority processing	Traffic policing	Queue scheduling
ONT	N/A	Trusts the network-layer 802.1p priority	(Optional) Limits downstream traffic bandwidth based on user-side ports.	PQ schedule based on 802.1p priority

15.14.2 FTTB/FTTC End-to-End QoS Policy

The FTTB/FTTC E2E QoS solution uses the DiffServ model. In this solution, QoS parameters are carried in each packet header to ensure QoS guarantees. The following describes upstream and downstream E2E QoS applications for FTTB/FTTC.

Upstream E2E QoS

Figure 15-66 shows the upstream E2E QoS.

1. QoS policy on the MDU.
   a. An MDU classifies traffic based on the user port or user-side VLAN IDs/PVC (xDSL access) to distinguish services
   b. Re-marks the 802.1p priority for service packets.
   c. Based on packet priorities, the MDU arranges packets into different priority queues by using PQ scheduling. This operation ensures that services with high QoS are forwarded preferentially.
   d. (Optional) Limits the upstream traffic on user-side ports.
2. QoS policy on the OLT.
   a. Classifies traffic based on VLAN IDs and the 802.1p priority
   b. Trusts the user-side priority or re-marks the 802.1p priority for service packets.
   c. Implements priority scheduling and congestion management based on the 802.1p priority.
   d. Implements DBA to limit the ONUs upstream bandwidth.
   e. (Optional) Limits upstream traffic bandwidth based on service flows.
3. The layer 2 metropolitan area network (MAN) between the OLT and the BRAS/service router (SR) implements priority scheduling based on the 802.1p priority of service packets.
4. Traffic classification is enabled on the ingress of the BRAS/SR to identify 802.1p values of different packets and re-mark packets according to the upper network.

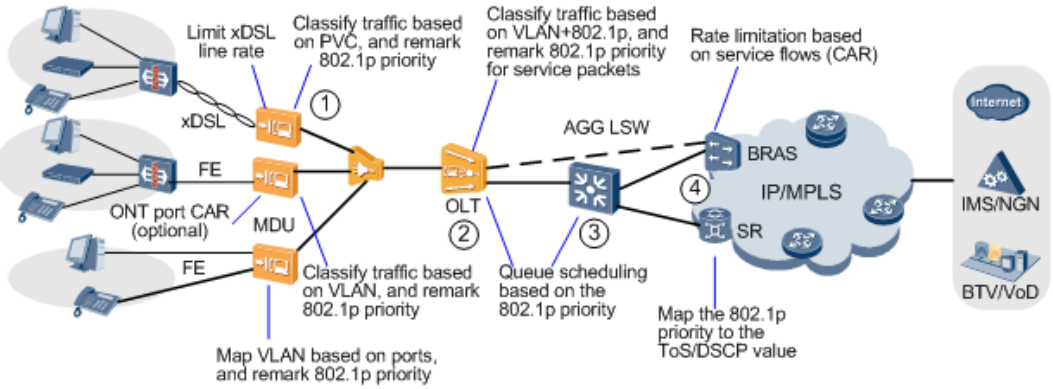
 Copyright © Huawei Technologies Co., Ltd.

- When the upper-layer network is a native IP Layer 3 network, the BRAS/SR maps the 802.1p priority to the ToS/DSCP value in the IP header.
- When the upper-layer network is an MPLS VPN, the BRAS/SR maps the 802.1p priority to MPLS EXP.

(Optional) The BRAS/SR can also limit bandwidth for upstream user traffic based on service flows

5. Upper-level networks of the BRAS/SR implement traffic classification, congestion management, and congestion avoidance based on ToS/DSCP in the IP header or the MPLS EXP priority.

Figure 15-66 FTTB/FTTC E2E QoS networking



Downstream E2E QoS

Figure 15-66 shows the downstream E2E QoS.

1. The packets of the IPTV, VoIP, and broadband Internet services are assigned different DSCP/MPLS EXP values based on priority requirements using related egress switches or provider edge (PE) devices. Core networks and backbone networks perform priority-based queue scheduling according to DSCP/MPLS EXP.
2. (Optional) Downstream packets are transmitted to edge BRAS/SR of the Layer 3 network through core networks and backbone networks. The BRAS/SR maps IP DSCP or MPLS EXP of packets to the VLAN 802.1p priority. The BRAS/SR can also limit bandwidth for downstream user traffic based on service flows.
   - For broadband Internet services in PPPoE mode, the BRAS controls the access bandwidth of each broadband user based on the rate authorized by the RADIUS server. For example, the rate can be 512 Kbit/s, 1 Mbit/s, or 2 Mbit/s.
   - Due to traffic jitter of the server or network, queue shaping is necessary for IPTV services on service routers.
3. Metropolitan networks and Layer 2 access networks downstream to the BRAS/SR implement priority scheduling based on the packet 802.1p priority.
4. (Optional) The OLT classifies traffic based on VLAN IDs and the 802.1p priority, and it trusts the network-side priority or re-marks the 802.1p priority for service packets. It implements priority scheduling and congestion

 Copyright © Huawei Technologies Co., Ltd.

management based on the 802.1p priority. It can limit downstream traffic bandwidth based on service flows.

5. (Optional) The MDU trusts the network-layer 802.1p priority and implements priority scheduling based on the 802.1p priority. The user-side port on the MDU limits bandwidth for downstream traffic.

15.14.3 QoS Solution for FTTH

Basic Principles of QoS for FTTH

1. The ingress (upstream port on ONT/downstream port on OLT) of an FTTx access network classifies traffic and re-marks the 802.1p priority for packets. The access device (MA5800) implements congestion control in PQ scheduling mode based on the 802.1p priority.
2. Service packets on an ONT are differentiated based on VLAN IDs.
   - In a GPON system, GEM port mapping is implemented based on VLAN IDs to specify whether different traffic packets are transmitted to different GEM ports or to the same GEM port. Each ONT uses a T-CONT.
3. DBA profiles are configured based on users' bandwidth types. During service provisioning, relevant DBA profiles are selected based on the bandwidth types that users apply for. The recommended DBA type is type 3 (assured bandwidth+maximum bandwidth). Upstream priority of the T-CONT for an ONT follows 802.1p-based priority scheduling.
4. It is recommended to use the same VLAN configuration for all ONTs and implement VLAN translation on the OLT. This operation ensures that the same type of ONTs having the same bandwidth types use the same line profile and service profile.
5. It is recommended to limit service flow bandwidth on the BRAS or service router (SR) but not on the access device.

   NOTE

   The ONT can be configured to limit bandwidth for user-side ports, and the OLT can be configured to rate limit service flows based on traffic profiles.

Recommended QoS Plan for FTTH

The following table provides details about the recommended QoS plan for FTTH with respect to traffic classification, priority marking, and queue scheduling policies.

Service Type	802.1p Priority	Queue Scheduling Method	OLT Queue ID (8 Queues)	ONT Queue ID	
				When Eight Queues Are Supported	When Four Queues Are Supported
Management service	6	PQ	6	6	3

 Copyright © Huawei Technologies Co., Ltd.

Service Type	802.1p Priority	Queue Scheduling Method	OLT Queue ID (8 Queues)	ONT Queue ID	
				When Eight Queues Are Supported	When Four Queues Are Supported
VoIP service	5	PQ	5	5	2
IPTV service	4	PQ	4	4	2
Internet access service	0	PQ	0	0	0

NOTE

The service priorities in this table are recommended values only. Configure the service priorities according to the operator's actual network plan.

The following table provides details about the recommended QoS plan for FTTH with respect to traffic policing and DBA policies.

Items	Management service	Internet access service	VoIP service	IPTV service
GEM port (only for GPON)	11 [Remark 1]	12	13	14
TCONT (only for GPON)	All services share a T-CONT.			
DBA type	Type 3 (DBA profile: assured bandwidth + maximum bandwidth. Users are allowed to preempt the bandwidth on condition that the users' assured bandwidth is guaranteed. However, the total bandwidth cannot exceed the maximum bandwidth.)			
DBA bandwidth planning	Configure the DBA bandwidth according to the user's bandwidth package. The assured bandwidth is the maximum bandwidth required by management packets, VoIP, and IPTV upstream packets. The maximum bandwidth is greater than or equal to the maximum bandwidth that users apply.			

 Copyright © Huawei Technologies Co., Ltd.

Items	Management service	Internet access service	VoIP service	IPTV service
Rate limit on OLT downstream	No rate limit	Configure rate limit by a traffic profile as required. [Remark 2]	No rate limit	No rate limit
Rate limit on ONU upstream port	No rate limit			
Rate limit on ONU downstream port	No rate limit			

Remark 1: GEM port values depend on the planning of the service provider. Generally, however, use different GEM ports for different services.

Remark 2: **Table 15-17** shows the reference service bandwidth of each service for each user.

Table 15-17 Reference service bandwidth of each service for each user

Service Type	Upstream Bandwidth	Downstream Bandwidth	Bandwidth Description
Internet access service	Determined based on the service package.	Determined based on the service package.	Available bandwidth of Internet access service = Committed bandwidth of the service package - VoIP bandwidth - IPTV bandwidth
VoIP service	200 Kbit/s	200 Kbit/s	The upstream bandwidth and the downstream bandwidth of VoIP service are symmetrical. The actual bandwidth is related to the coding format. This bandwidth is calculated for two POTS ports.

 Copyright © Huawei Technologies Co., Ltd.

Service Type	Upstream Bandwidth	Downstream Bandwidth	Bandwidth Description
IPTV service (standard definition program)	—	2.5 Mbit/s per channel	IPTV service mainly occupies the downstream bandwidth. The actual bandwidth depends on the coding format, the picture in picture (PiP) information, 10% of the assured bandwidth for burst traffic, and the number of programs that can be concurrently watched by one user (in the case of multiple STBs). The upstream bandwidth is mainly used for transmitting IGMP packets, which requires only a little bandwidth. Therefore, the bandwidth occupied by IGMP packets can be ignored.
IPTV service (high definition program)	—	9.7 Mbit/s per channel	

Note:

- If the BRAS does not support rate limitation, OLTs can limit the rate for service flows by using traffic profiles.
- Different service packets on the ONT are distinguished by different VLAN IDs. The service packets are mapped to GEM ports based on VLAN IDs so that different service packets are transmitted to different GEM ports. Each GEM port (each service) can correspond to a T-CONT or all GEM ports share a T-CONT.
- The sum of the assured bandwidth of all ONTs connected to an OLT PON port and the fixed bandwidth of OMCI management channel should be smaller than the GPON upstream bandwidth. Some bandwidth must be reserved for future service expansion.

15.14.4 QoS Solution for FTTB/FTTC

Basic Principles of QoS for FTTB/FTTC

1. Plan different CoS priorities for different services. The advised priority order is as follows: management packets > VoIP > IPTV > HSI.
2. Classify traffic and re-mark 802.1p priority on the ingress (upstream port on MDU/downstream port on OLT) on the FTTx network. Implement congestion control based on 802.1p priority in PQ scheduling mode on the internal access network (MDU/OLT). For packets transmitted downstream from the network side of the OLT, the priority can be trusted. The priority does not need to be re-marked.
3. Service packets on MDUs are distinguished by 802.1p priority.

 Copyright © Huawei Technologies Co., Ltd.

- GEM port mapping is implemented in a GPON system based on 802.1p, so that different service packets enter different GEM ports. Each GEM port (each service) can correspond to a T-CONT or all GEM ports share a T-CONT.

  NOTE

  T-CONT 0 is still used exclusively for OMCI management.

4. DBA planning:
   - GPON MDUs sharing one T-CONT: Set different DBA profiles for different MDU types on the OLT, and then choose corresponding DBA profiles in the service pre-provisioning phase based on the MDU type. Type 3 (assured bandwidth+maximum bandwidth) is recommended for a DBA profile. The T-CONT upstream priority for MDUs is based on 802.1p priority.
   - GPON MDUs using multiple T-CONTs: Allocate different DBA profiles for T-CONTs of each service.
5. Bandwidth limit: It is recommended to implement bandwidth limit of user service flows on the BRAS/SR, not on the access network. MDUs can be configured to limit the bandwidth on the user-side ports; MDUs/OLTs can be configured to limit the rate of service flows based on traffic profiles.
6. For xDSL access, an MxU supports bandwidth limit on physical links using the xDSL link traffic profile. Bandwidth limit on xDSL link is based on the fact that the channels on user-side ports do not differentiate multiple service PVCs. It is recommended that you differentiate services on the modem for users with large upstream traffic so that QoS can be ensured.

Recommended QoS Plan for FTTB/FTTC

The following table provides details about the recommended QoS plan for FTTB/FTTC with respect to traffic classification, priority marking, and queue scheduling policies.

Service Type	802.1p Priority	Queue Scheduling Method	OLT Queue ID (8 Queues)	MDU Queue ID	
				When Eight Queues Are Supported	When Four Queues Are Supported
Management service	6	PQ	6	6	3
VoIP service	5	PQ	5	5	2
IPTV service	4	PQ	4	4	2
Internet access service	0	PQ	0	0	0

 Copyright © Huawei Technologies Co., Ltd.

Note

- Different service packets are distinguished by different VLAN IDs. GEM ports are mapped based on 802.1p priorities for the GPON system.
- Service priorities in this table are for reference only. Configure the service priorities according to the operator's actual network plan.

The following table provides details about the recommended QoS plan for FTTB/FTTC with respect to traffic policing and DBA policies.

Items	Management service	Internet access service	VoIP service	IPTV service
GEM port (only for GPON)	11 [Remark 1]	12	13	14
T-CONT (only for GPON)	All services share a T-CONT.			
DBA type	Type 3 (DBA profile: assured bandwidth + maximum bandwidth. Users are allowed to preempt the bandwidth on condition that the users' assured bandwidth is guaranteed. However, the total bandwidth cannot exceed the maximum bandwidth.)			
DBA bandwidth planning	Configure the DBA bandwidth according to the user's bandwidth package. The assured bandwidth is the maximum bandwidth required by management packets, VoIP, and IPTV upstream packets. The maximum bandwidth is greater than or equal to the maximum bandwidth that users apply.			
Rate limit on OLT downstream	No rate limit	Configure rate limit by a traffic profile as required. [Remark 2]	No rate limit	No rate limit
Rate limit on ONU upstream port	Set ONU port rate limit or xDSL line rate limit as required. [Remark 2]			
Rate limit on ONU downstream port	Set ONU port rate limit or xDSL line rate limit as required. [Remark 2]			

Remark 1: GEM port values depend on the planning of the service provider. Generally, however, use different GEM ports for different services.

Remark 2: **Table 15-18** shows the reference service bandwidth of each service for each user.

 Copyright © Huawei Technologies Co., Ltd.

Table 15-18 Reference service bandwidth of each service for each user

Service Type	Upstream Bandwidth	Downstream Bandwidth	Bandwidth Description
Internet access service	Determined based on the service package.	Determined based on the service package.	Available bandwidth of Internet access service = Committed bandwidth of the service package - VoIP bandwidth - IPTV bandwidth
VoIP service	200 Kbit/s	200 Kbit/s	The upstream bandwidth and the downstream bandwidth of VoIP service are symmetrical. The actual bandwidth varies with the coding format. This bandwidth is calculated for two POTS ports.
IPTV service (common program)	/	2.5 Mbit/s per channel	IPTV service mainly occupies the downstream bandwidth. The actual bandwidth depends on the coding format, the picture in picture (PiP) information, 10% of the assured bandwidth for burst traffic, and the number of programs that can be concurrently watched by one user (in the case of multiple STBs). The upstream bandwidth is mainly used for transmitting IGMP packets, which requires a little bandwidth and can be ignored.
IPTV service (high definition program)	/	9.7 Mbit/s per channel	

Note

- It is recommended to configure rate limitation on the BRAS or SR, not on the OLTs or MDUs. If the BRAS does not support rate limit, OLTs can limit rates on service flows using traffic profiles.
- Different service packets of MDUs are distinguished by 802.1p priorities and are mapped to GEM ports based on 802.1p priorities so that packets are transmitted to different GEM ports. Each GEM port (each service) can correspond to a T-CONT or all GEM ports share a T-CONT.
- The sum of the assured bandwidth of all ONTs connected to an OLT PON port and the fixed bandwidth of OMCI management channel should be smaller than the GPON upstream bandwidth. Some bandwidth must be reserved for future service expansion.

 Copyright © Huawei Technologies Co., Ltd.

16 Reliability

This topic describes reliability features implemented by the system.

16.1 Reliability Solution Overview
The OLT provides a comprehensive reliability solution, including uplink protection, device protection, and PON downlink protection, enhancing reliability of the entire network.

16.2 xPON Type B Protection
xPON type B protection allows dual-channel redundancy protection for OLT PON ports and backbone fibers on a xPON network. This feature improves ODN network reliability and ensures service continuity.

16.3 xPON Type C Protection
The xPON type C protection switching is implemented through the redundancy configuration of the two PON ports on the ONU, backbone optical fiber, optical splitter, and tributary optical fiber on a xPON network. Each item is in a dual configuration. The protection improves the reliability on the optical distribution network (ODN) and prevents service interruption.

16.4 Redundancy Backup of Control Boards
An OLT configured with two control boards works in the load sharing mode. When working normally, the active and standby control boards concurrently forward service packets.

16.5 Ethernet Port Protection Group
An Ethernet port protection group binds two or more Ethernet ports together. The working port carries services, while the protection port provides backup and does not carry services. If the working port is faulty, the services carried by it are switched to the protection port, automatically or manually as required. Ethernet port protection ensures uninterrupted service transmission and improves reliability of links.

16.6 Ethernet Link Aggregation
Ethernet link aggregation is a process that aggregates two or more Ethernet ports of the same type to a logic port. It increases the link bandwidth without requiring a hardware upgrade and improves link reliability using the link backup mechanism.

16.7 MSTP
The Multiple Spanning Tree Protocol (MSTP) is compatible with STP and RSTP.

 Copyright © Huawei Technologies Co., Ltd.

16.8 ERPS
Ethernet ring protection switching (ERPS) is a ring network protocol defined in Recommendation ITU-T G.8032. Interoperation is achieved if all devices participating in the ring network support ERPS. On a network composed of ERPS-supporting devices, the service protection solution with quick switching can be implemented using a small number of links.

16.9 VRRP Snooping
VRRP is a fault-tolerant protocol. It allows multiple routers to form a virtual routing device, and provides a mechanism, which ensures that services will be taken over in time by another device once the next hop of a host fails. In this way the continuity and reliability of communication are ensured. VRRP snooping is to snoop (or listen for) VRRP packets.

16.10 Ring Check
The ring check feature is mainly used to detect the user-side, cascade-side, and network-side ring networks, and eliminate the user-side ring network.

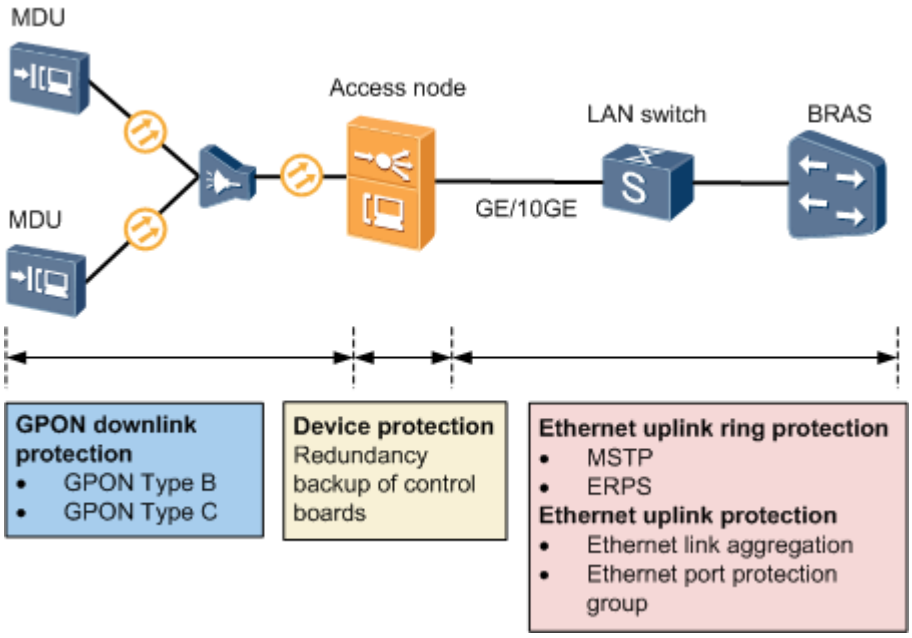
16.11 Monitor Link
A monitor link group is composed of one uplink and several downlinks. It is used to monitor the fault of uplinks.

16.1 Reliability Solution Overview

The OLT provides a comprehensive reliability solution, including uplink protection, device protection, and PON downlink protection, enhancing reliability of the entire network.

To ensure reliability of the access network, the OLT supports the following end-to-end (E2E) reliability solution.

Figure 16-1 Reliability scheme in an access network



 Copyright © Huawei Technologies Co., Ltd.

Figure 16-2 Reliability scheme in an access network

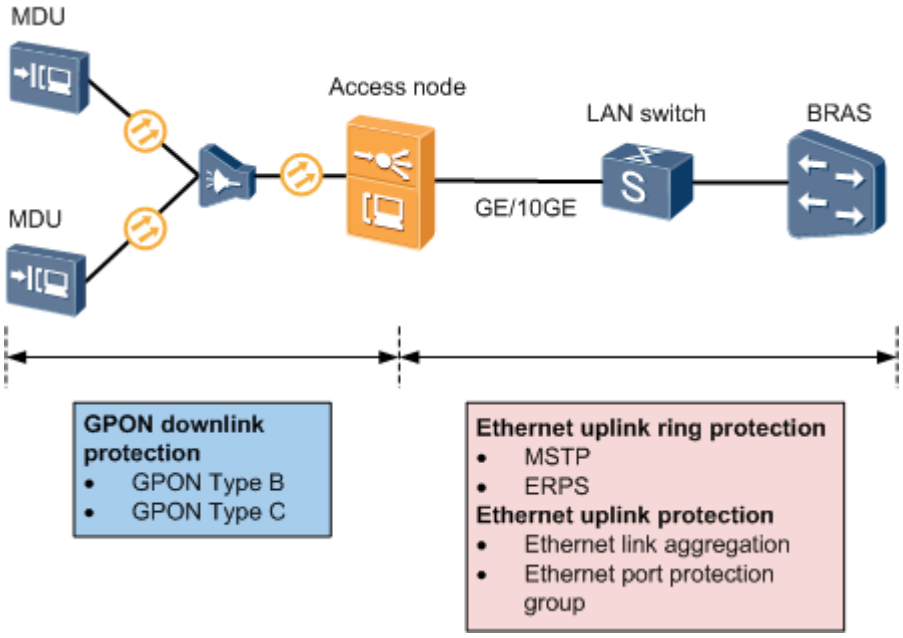


Table 16-1 Reliability scheme

Protected Object	Protection Scheme	Huawei Proprietary Protocol or Not	Function and Feature
Device hardware	Redundancy backup of control boards	No	An Ethernet port protection group can be configured only after active/standby redundancy backup of control boards is configured. The load-sharing mode doubles bandwidths and improves data forwarding performance of the device.
Ethernet uplinks or cascade links	Ethernet link aggregation • Manual aggregation • LACP aggregation	No	Increases uplink bandwidths, achieving load sharing. Members in the aggregation group back up for each other, enhancing uplink reliability. Manual aggregation is required if the OLT is connected to a device that does not support LACP.
	Ethernet port protection group	No	Supports uplink port protection, which can be used with Ethernet link aggregation.

 Copyright © Huawei Technologies Co., Ltd.

Protected Object	Protection Scheme	Huawei Proprietary Protocol or Not	Function and Feature
Ethernet uplink ring protection	MSTP	No	MSTP prunes a loop network to a loop-free tree network to avoid proliferation and infinite loop of packets in the loop network. MSTP is compatible with STP and RSTP. Furthermore, MSTP remedies drawbacks of STP and RSTP. The convergence time is subject to the network topology.
	ERPS	No	ERPS is a ring network protocol defined in Recommendation ITU-T G.8032. Interoperation is achieved if all devices participating in the ring network support ERPS. In addition, ERPS supports fast convergence, meeting carrier-class reliability requirements.
GPON downlink protection	GPON type B	No	Type B provides redundancy for OLT's GPON ports and feeder fiber. Compared with type C, type B requires a lower cost but the implementation is more difficult. **NOTE** When the number of ONTs/ONUs connected to a board exceeds 1000, you are advised to configure inter-board Type B protection.
	GPON type C	Yes	Type C provides redundancy for ONU's GPON ports, feeder fibers, splitters, and distribution fibers. Compared with type B, type C provides higher reliability.

16.2 xPON Type B Protection

xPON type B protection allows dual-channel redundancy protection for OLT PON ports and backbone fibers on a xPON network. This feature improves ODN network reliability and ensures service continuity.

16.2.1 Introduction to xPON Type B Protection

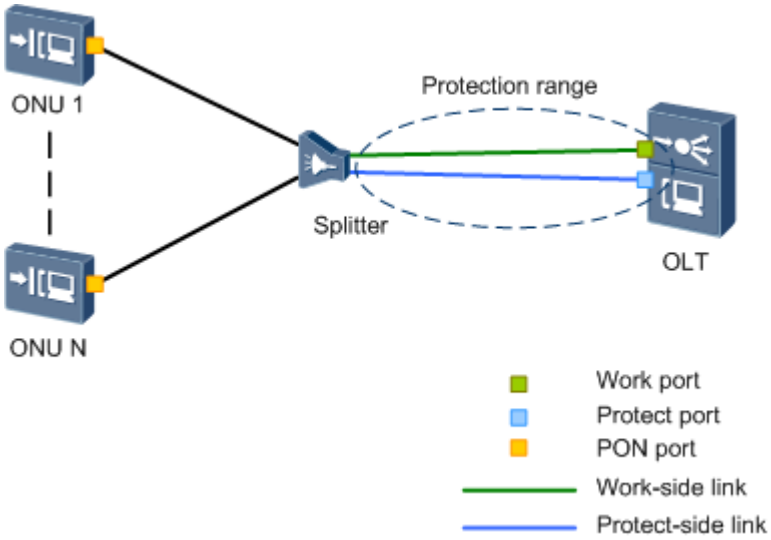
Type B provides protection for the OLT, OLT PON ports, and backbone fiber.

 Copyright © Huawei Technologies Co., Ltd.

xPON type B applies to single-homing or dual-homing scenarios.

Figure 16-3 shows a single-homed xPON type B protection network. Type B single-homing protection mainly protects the OLT PON port and backbone fibers. When the OLT PON port or the backbone fiber is faulty, services can be automatically switched to another OLT PON port or the backbone fiber.

Figure 16-3 Single-homed xPON type B protection network

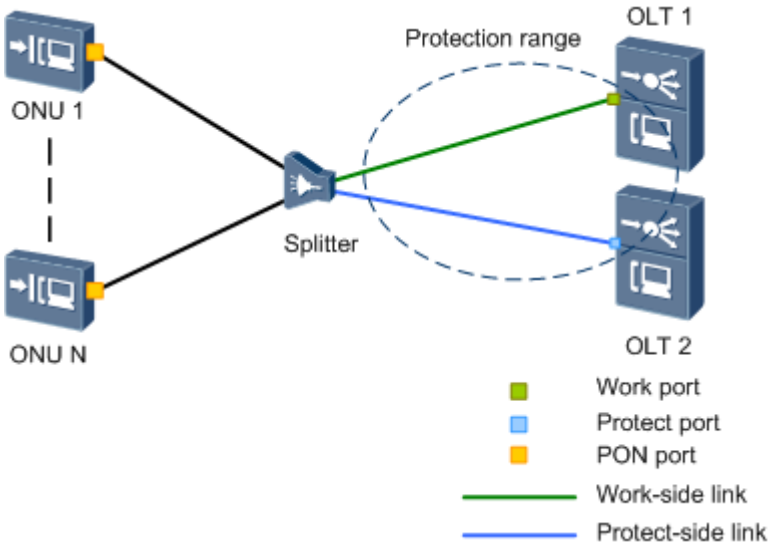


Networking Scenario	Advantage	Disadvantage
Single homing	Networking, OLT/ONU management, and service provisioning are simple.	An OLT fault will interrupt services. In addition, two optical fibers routed in one pipe may both be broken.

Figure 16-4 shows a dual-homed xPON type B protection network. Type B dual-homing protection mainly protects the OLT, OLT PON port, and backbone fibers. When the OLT, OLT PON port, or backbone fiber is faulty, services can be automatically switched to another OLT, OLT PON port, or backbone fiber.

 Copyright © Huawei Technologies Co., Ltd.

Figure 16-4 Dual-homed xPON type B protection network



Networking Scenario	Advantage	Disadvantage
Dual homing	Each of the two OLTs connects to a backbone fiber for remote disaster recovery.	The networking is complex, and networking costs are high. In addition, the OLT configuration is complex.

16.2.2 Basic Concepts of xPON Type B Protection

Protection Group

On a single-homed network, two PON ports on an OLT are added to a protection group.

NOTE

The OLT PON ports can be on the same board or on different boards. The differences are as follows:

- Port redundancy backup on the same board can conserve hardware resources. If the PON service board fails, the services on the entire board are interrupted.
- Port redundancy backup on the different boards requires hardware costs than that on the same board. If the active PON service board fails, the services can be automatically switched over to the PON ports on the standby board without being interrupted.

On a dual-homed network, the PON ports on two OLTs are added to a protection group.

Roles of Protection Group Members

Protection group members have two roles: working and protection.

 Copyright © Huawei Technologies Co., Ltd.

- One protection group contains a working port and a protection port. The working port and protection port are two different PON ports.
- In normal cases, the working port carries services.
- When the link of the working port becomes faulty, the system automatically switches services from the working port to the protection port to ensure service continuity.

State of Protection Group Members

Protection group members have two states: active and standby. The active port forwards data and the standby port does not forward data.

Switching Types

The switching can be triggered automatically by a fault or performed manually. Manual operations that may cause switching are forcible switching and locking.

- In automatic switching, the OLT and ONU automatically switch to the standby link when the conditions for triggering the switching are met.
- In forcible switching, users run the **force-switch** command on the OLT to perform the switching regardless of whether specific group members are running properly.
- After switching, the working port's state becomes standby. Then, if users run the **lockout** command on the OLT to lock a group member (for a Type B single-homed network, both working and protection ports can be locked; for a Type B dual-homed network, only the protection port can be locked), the switching is performed and the working port's state becomes active.

Switching Condition

Protection group members are switched only when the following conditions are met:

- The protection group is enabled.

  NOTE

  The status of a protection group can be queried using the **display protect-group** command on the OLT. If **Admin State** is displayed in the output, the protection group is enabled.
- The protection group is not frozen using the **freeze** command on the OLT.
- The protection group is not locked using the **lockout** command on the OLT.
- The protection group member is not forcibly switched using the **force-switch** command on the OLT.

 Copyright © Huawei Technologies Co., Ltd.

Operation Restriction Relationships in Protection Switching

Table 16-2 Type B single homing protection switching

Current Status				Next Operation									Remarks
Enabled	Frozen	Locked	Forcible switching	Enabled	Disabled	Frozen	Unfrozen	Locked	Unlocked	Forcible switching	Canceling forcible switching	Automatic switching	None
No	No	No	No	Supported	N/A	N/A	N/A	Supported	N/A	N/A	N/A	N/A	None
No	No	Yes	No	Supported	N/A	N/A	N/A	N/A	Supported	N/A	N/A	N/A	None
Yes	No	No	No	N/A	Supported	Supported	N/A	Supported	N/A	Supported	N/A	Supported	None

 Copyright © Huawei Technologies Co., Ltd.

Current Status				Next Operation									Remarks
Enabled	Frozen	Locked	Forcible switching	Enabled	Disabled	Frozen	Unfrozen	Locked	Unlocked	Forcible switching	Canceling forcible switching	Automatic switching	None
Yes	No	No	Yes	N/A	Supported	Supported	N/A	Supported	N/A	Supported	N/A	N/A	The forcible switching status will be cleared when the protection switching is disabled or locked.
Yes	No	Yes	No	N/A	Supported	Supported	N/A	N/A	Supported	N/A	N/A	N/A	None

Current Status				Next Operation									Remarks
Enabled	Frozen	Locked	Forcible switching	Enabled	Disabled	Frozen	Unfrozen	Locked	Unlocked	Forcible switching	Canceling forcible switching	Automatic switching	None
Yes	Yes	No	No	N/A	N/A	N/A	Supported	N/A	N/A	N/A	N/A	N/A	None
Yes	Yes	No	Yes	N/A	N/A	N/A	Supported	N/A	N/A	N/A	N/A	N/A	None
Yes	Yes	Yes	No	N/A	N/A	N/A	Supported	N/A	N/A	N/A	N/A	N/A	None

NOTE

The statuses that are not listed in the preceding table, such as disabled and forcible switching, are unavailable.

Table 16-3 Type B dual homing protection switching

Current Status			Next Operation							Remarks
Enabled	Locked	Forcible switching	Enabled	Disabled	Locked	Unlocked	Forcible switching	Canceling forcible switching	Automatic switching	None

Current Status			Next Operation							Remarks
No	No	No	Supported	N/A	Supported	N/A	N/A	N/A	N/A	Protection switching is consistently enabled on the working side, and the supported statuses are available only on the protection side.

 Copyright © Huawei Technologies Co., Ltd.

Current Status			Next Operation						Remarks
No	Yes	No	Supported	N/A	N/A	Supported	N/A	N/A	N/A	Protection switching is consistently enabled on the working side, and the supported statuses are available only on the protection side.

 Copyright © Huawei Technologies Co., Ltd.

Current Status			Next Operation							Remarks
Yes	No	No	N/A	Supported	Supported	N/A	Supported	N/A	Supported	Protection switching can be disabled or locked only on the protection side.

Current Status			Next Operation							Remarks
Yes	No	Yes	N/A	Supported	Supported	N/A	Supported	Supported	N/A	Protection switching can be disabled or locked only on the protection side. In either of the statuses, the forcible switching status will be cleared.

 Copyright © Huawei Technologies Co., Ltd.

Current Status			Next Operation							Remarks
Yes	Yes	No	N/A	Supported	N/A	Supported	N/A	N/A	N/A	Protection switching can be locked only on the protection side.

NOTE

The statuses that are not listed in the preceding table, such as disabled and forcible switching, are unavailable.

Associated Switching

Associated switching is implemented on a dual-homed network as follows: A protection group is associated on the OLT with the uplink Ethernet port status and BFD/MEP session status. In such a case, when the OLT's upper-layer network or the Layer 2 OLT physical link fails, the active OLT triggers a dual-homing protection switchover so that services will be switched to the standby OLT.

16.2.3 Single-Homing xPON Type B Protection Principles

On a single-homed network, the two PON ports on the OLT are in active/standby state, and they cannot forward packets at the same time. An automatic switchover can be triggered by any of the following conditions:

- Active fiber cut
- Active PON port failure
- Line quality deterioration

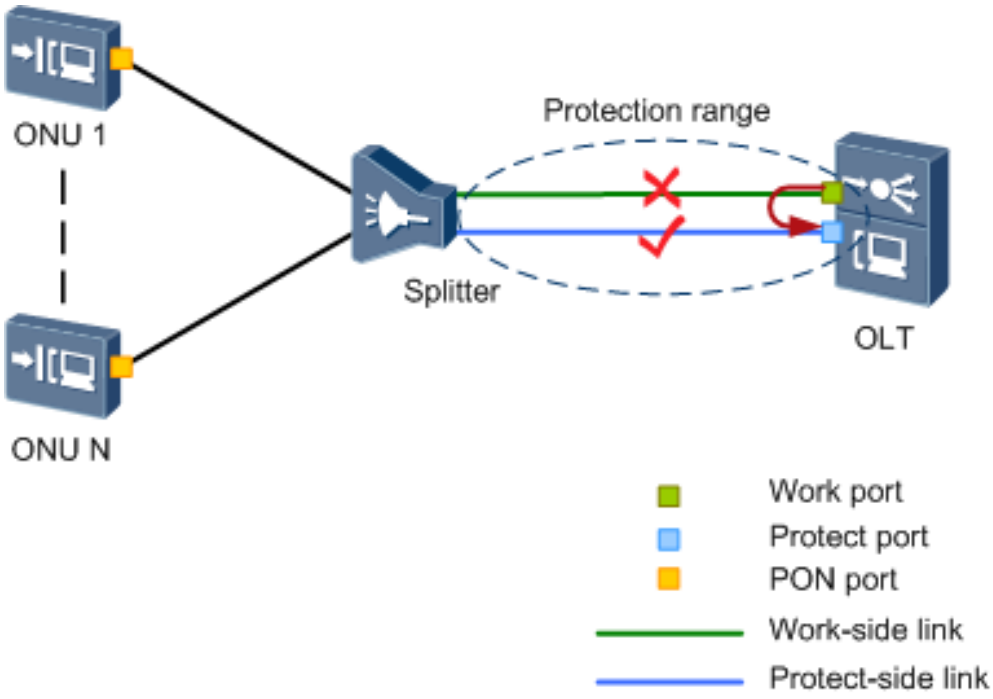
The following section describes the xPON type B protection switching in different scenarios.

Scenario 1: Active Optical Fiber Is Cut

The active optical fiber is cut when the OLT PON port is working, as shown in **Figure 16-5**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 16-5 Active optical fiber is cut



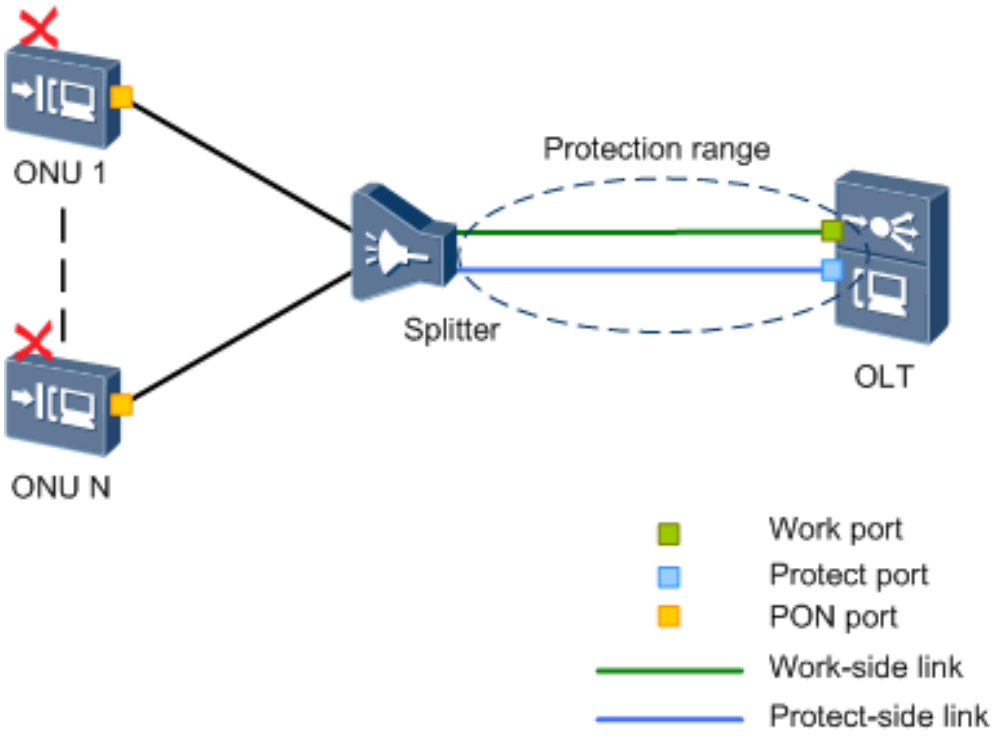
- When detecting a loss of signal (LOS) alarm (generated due to the active optical fiber cut), the working port disables the transmission of the optical module.
- The protection port detects a LOS alarm (generated due to the failure of the ONT to transmit optical signals), enables the transmission function of the optical module, and performs ONU ranging.
- If the optical fiber connected to the protection port is functional, and ONU ranging is successful, the protection port reports an LOS clear alarm.
- The working port switches to the standby state. The protection port switches to the active state. Then, the protection switching ends.

Scenario 2: All ONUs Go Offline

All ONUs connected to the OLT PON port go offline, as shown in **Figure 16-6**.

Figure 16-6 All ONUs go offline

 Copyright © Huawei Technologies Co., Ltd.



- When detecting an LOS alarm (generated because all ONUs go offline), the working port disables the transmission of the optical module.
- The protection port detects a LOS alarm (generated due to the failure of the ONT to transmit optical signals), enables the transmission function of the optical module, and performs ONU ranging.
- No ONU connected to a PON port goes online due to a ranging failure. Therefore, the OLT cyclically detects the working and protection ports until an ONU goes online.
- After the ONU goes online, switching is performed between the PON ports if the protection port is detected. If the protection port is not detected, the working port continues working.

16.2.4 Dual-Homing xPON Type B Protection Principles

On a dual-homed network, two OLTs protection work in active/standby mode and they cannot forward packets at the same time. An automatic switchover can be triggered by any of the following conditions:

NOTE

Currently, type B dual-homing protection is mainly used in the passive optical LAN (POL) solution and requires cooperation of the OLT, ONU, and aggregation device. You are advised to set up a cluster or VRRP interconnection on the core switches to work with type B dual-homing protection to enhance the overall reliability of the solution.

- Optical fiber cut from the active OLT
- Active OLT fault
- Active OLT PON port fault

 Copyright © Huawei Technologies Co., Ltd.

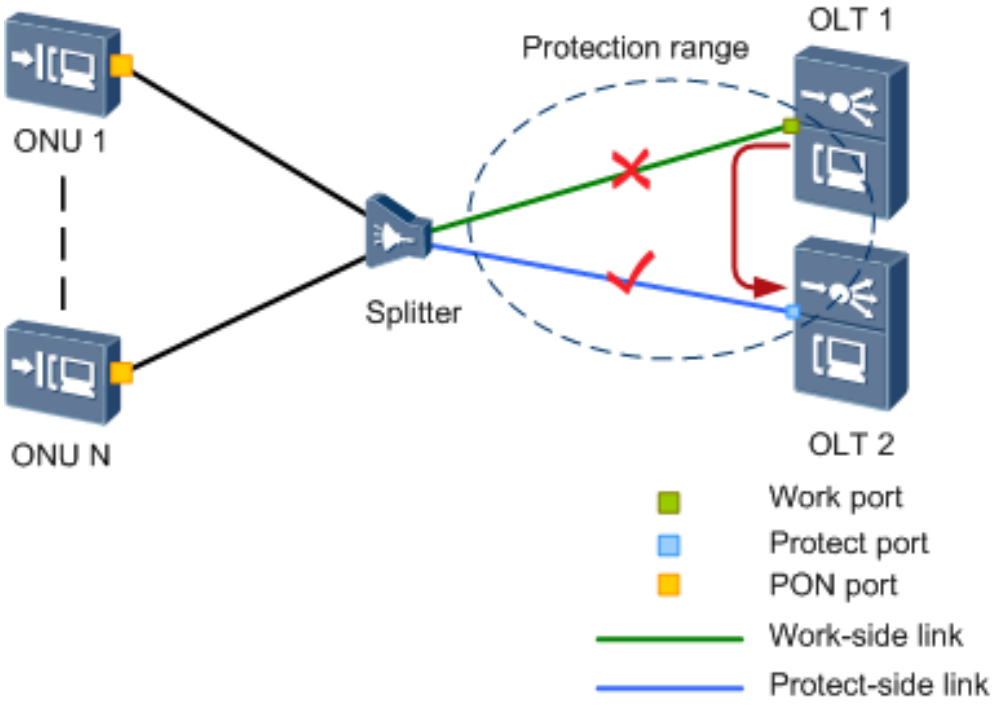
- The OLT's uplink is faulty (this condition triggers automatic switching only in the associated protection switching scenario).
- Active line quality deterioration

The following section describes the xPON type B protection switching in different scenarios.

Scenario 1: Active Optical Fiber Is Cut

The active optical fiber is cut when the OLT PON port is working, as shown in **Figure 16-7**.

Figure 16-7 Active optical fiber is cut



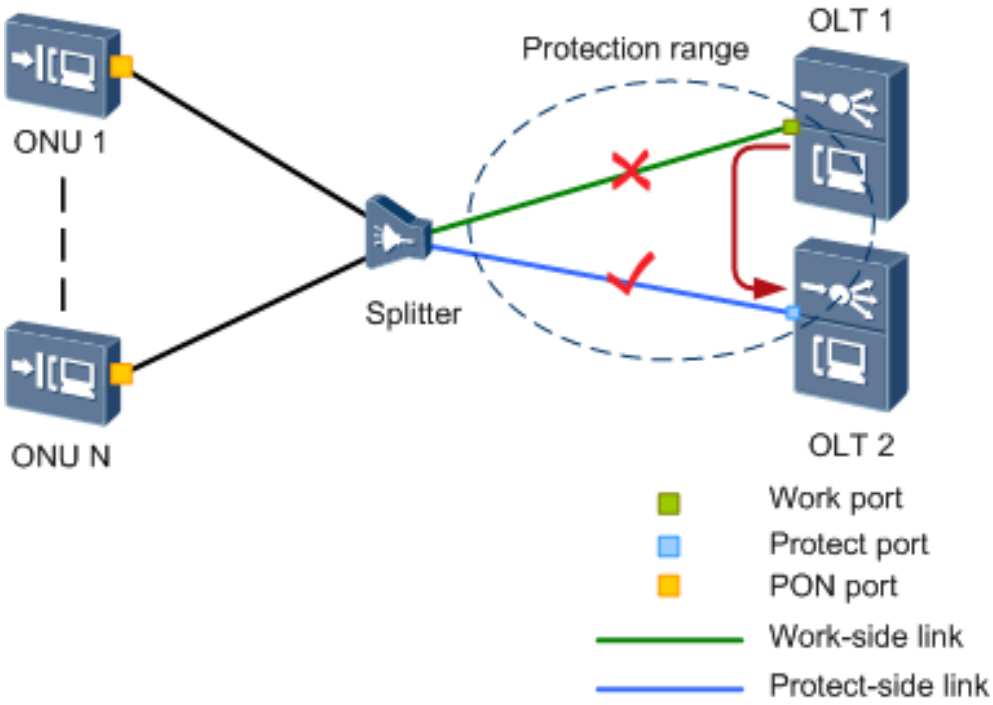
- When detecting a loss of signal (LOS) alarm (generated due to the active optical fiber cut), the working port disables the transmission of the optical module.
- The protection port detects a LOS alarm (generated due to the failure of the ONT to transmit optical signals), enables the transmission function of the optical module, and performs ONU ranging.
- If the optical fiber connected to the protection port is functional, and ONU ranging is successful, the protection port reports an LOS clear alarm.
- The working port switches to the standby state. The protection port switches to the active state. Then, the protection switching ends.

 Copyright © Huawei Technologies Co., Ltd.

Scenario 2: All ONUs Go Offline

All ONUs connected to the OLT PON port go offline, as shown in **Figure 16-8**.

Figure 16-8 All ONUs go offline



- When detecting an LOS alarm (generated because all ONUs go offline), the working port disables the transmission of the optical module.
- The protection port detects a LOS alarm (generated due to the failure of the ONT to transmit optical signals), enables the transmission function of the optical module, and performs ONU ranging.
- No ONU connected to a PON port goes online due to a ranging failure. Therefore, the OLT cyclically detects the working and protection ports until an ONU goes online.
- After the ONU goes online, switching is performed between the PON ports if the protection port is detected. If the protection port is not detected, the working port continues working.

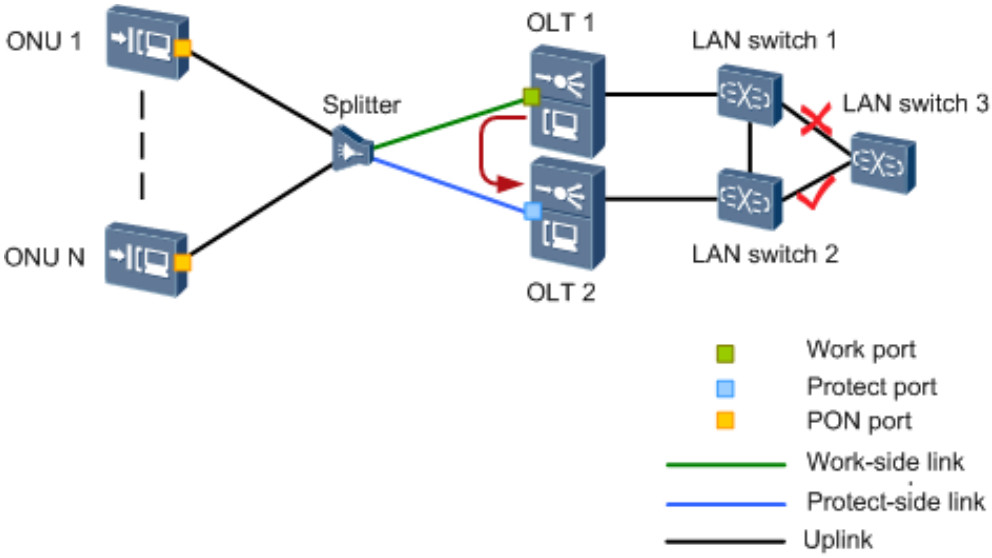
Scenario 3: Associated Protection Switching Is Caused by a Connection Failure in the OLT's Upstream Transmission Network

If a BFD session has been configured on the OLT, the BFD session can be bound to a protection group for creating an association between them. If CFM has been enabled on an OLT, an MEP session can be bound to a protection group for creating an association between them. Based on the associations, when the upper-layer network connection of the active OLT fails, the active and standby OLTs

 Copyright © Huawei Technologies Co., Ltd.

perform a switchover and notifies the ONU of the switchover. In this way, services are restored. **Figure 16-9** shows the associated protection switching caused by a connection failure in the OLT's upstream transmission network.

Figure 16-9 OLT's upstream transmission network connection fails



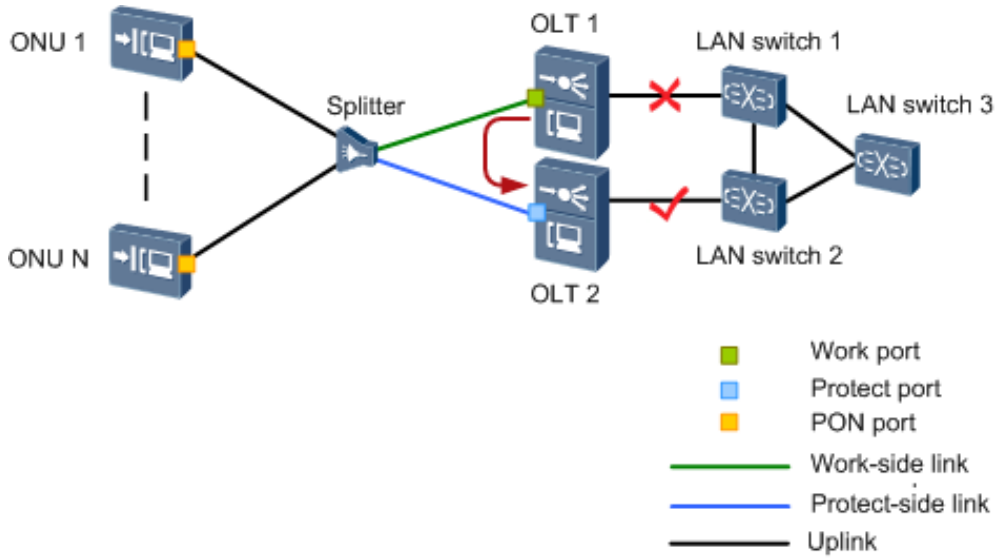
- On the OLT, the dual-homing protection group is associated with the BFD or MEP session. When the upstream route of the working OLT (OLT 1) fails, OLT 1 checks whether the protection OLT (OLT 2) and the upstream route of OLT 2 are functional. If they are functional, and both OLTs do not carry out a forcible switchover or locking operation, OLT 1 data has been synchronized to OLT 2. Then, the two OLTs perform a switchover.
- OLT 2 starts to work. It enables the transmission of the optical module and performs ONU ranging.
- After the switchover, the ONU service data is sent to OLT 2 through the protection port, and service data is transmitted over the protection link. The OLT status changes as follows:
  - OLT 1 becomes standby.
  - OLT 2 becomes active.

Scenario 4: Associated Protection Switching Is Caused by an OLT's Layer 2 Physical Link Fault

An OLT protection group is associated with the uplink Ethernet port status. Based on the association, when the Layer 2 physical link of the active OLT fails, the two OLTs perform a switchover, and notify the ONU of the switchover. In this way, services are restored. **Figure 16-10** shows the associated protection switching caused by an OLT's Layer 2 physical link fault.

Figure 16-10 OLT's Layer 2 physical link fails

 Copyright © Huawei Technologies Co., Ltd.



- On the OLT, the dual-homing protection group is associated with the uplink Ethernet port status. When the Ethernet port associated with the protection group on the working OLT (OLT 1) becomes **Down**, OLT 1 checks whether the protection OLT (OLT 2) and the physical link of the protection OLT are functional. If they are functional, and both OLTs do not carry out a forcible switchover or locking operation, OLT 1 data has been synchronized to OLT 2. Then, the two OLTs perform a switchover.
- OLT 2 starts to work. It enables the transmission of the optical module and performs ONU ranging.
- After the switchover, the ONU service data is sent to OLT 2 through the protection port, and service data is transmitted over the protection link. The OLT status changes as follows:
  - OLT 1 becomes standby.
  - OLT 2 becomes active.

16.2.5 Configuring xPON Type B Single-Homing Protection

This section describes how to configure xPON type B single-homing protection on an OLT to implement xPON port 1+1 redundancy backup, which ensures that services are not interrupted if a fault occurs on the OLT's PON port or backbone fiber.

Precaution

After xPON type B single-homing protection is configured, the service configurations on an optical network unit (ONU) remain unchanged and data is transmitted or received over the primary xPON port.

NOTE

Configurations of xPON network protection do not affect service configuration. This topic describes how to perform configurations of xPON network protection. For details about how to perform service configuration, see the specific service configuration guide.

 Copyright © Huawei Technologies Co., Ltd.

Procedure

Step 1 (Optional)Shutdown the laser of the protection member PON port.

NOTE

This step applies to the following scenarios, Otherwise, skip this step.

- The ONU upstream port works in auto-negotiation mode. That is, the ONU upstream port can work in GPON mode or XG(S)-PON mode according to the port type on the OLT side.
- The optical fiber of the protection member PON port of Type X protection is connected, that is, the physical link is connected. In this case, the physical link is connected but the Type X protection data configuration is not complete. The working and protection member PON ports transmit light at the same time and interact with the ONU. As a result, the upstream port mode of the ONU is repeatedly switched and the ONU is repeatedly reset.

To solve the preceding problem:

1. Run the **shutdown** command to disable the laser on the protection member PON port before the configuration.
2. After the configuration is complete, run the **undo shutdown** command to turn on the laser of the protection member PON port.

Step 2 Run the **protect-group** command to create a protection group for xPON access ports.

NOTE

1. Set **protect-target** to **gpon-uni-port** when GPON Type B single-homing protection is configured.
2. The working mode of the members in the protection group can only be **timedelay**.

Step 3 Run the **protect-group member** command to add a protection member to the protection group.

Step 4 Run the **protect-group enable** command to enable the protection group.

A created protection group is disabled by default. Enable the protection group to make the protection group take effect.

Step 5 Run the **display protect-group** command to query the information about the protection group and all the members in the protection group.

Step 6 Bind a PPPoE single MAC address pool to a protection group if PPPoE single MAC is enabled.

- To do so, run the **bind mac-pool single-mac** command in the protect-group mode.
- Otherwise, the PPPoE service carried over the xPON port is interrupted when a protection switchover is performed. In this case, users need to dial numbers up again to go online and the service interruption time is based on BRAS configurations. This may fail to meet the switchover performance requirements of no longer than 50 ms for a protection switchover.

Step 7 (Optional)Turn on the laser of the protection member PON port. After the preceding configuration steps are complete, run the **undo shutdown** command to turn on the laser of the protection member PON port.

----End

 Copyright © Huawei Technologies Co., Ltd.

Result

After the configuration, the primary xPON port on the OLT works in active mode and the secondary xPON port works in standby mode.

An automatic switching can be triggered by any of the following conditions:

- Fractures of optical fibers
- OLT xPON port failure
- Quality deterioration of lines

For a type B inter-board protection group, if the customer needs to change the board that houses the active PON port, trigger type B protection switching to activate the standby PON port, and then migrate the to-be-migrated PON port to the target board.

Example

The following configurations are used as an example to configure GPON type B single-homing protection on the OLT:

NOTE

The data in the example is for reference only. Replace it with the service data in the actual configuration.

- Protection group members: 0/2/0 and 0/2/1 (on the same GPON service board)
- Primary port: 0/2/0
- Secondary port: 0/2/1

```
huawei(config)#interface gpon 0/2
huawei(config-if-gpon-0/2)#shutdown 1
huawei(config-if-gpon-0/2)#quit
huawei(config)#protect-group 0 protect-target gpon-uni-port workmode timedelay
huawei(protect-group-0)#protect-group member port 0/2/0 role work
huawei(protect-group-0)#protect-group member port 0/2/1 role protect
huawei(protect-group-0)#protect-group enable
huawei(config)#interface gpon 0/2
huawei(config-if-gpon-0/2)#undo shutdown 1
huawei(config-if-gpon-0/2)#quit
huawei(config)#save
```

16.2.6 Configuring xPON Type B Dual-Homing Protection

In xPON type B dual-homing protection, 2 backbone fibers connect to 2 OLTs to realize remote disaster recovery.

Prerequisites

An ONU has been added to the active and standby OLTs by running the **ont add** command. All ONU profiles, such as DBA profile and line profiles, are the same on the active and standby OLTs. When adding an ONU to an OLT, ensure that the ONU IDs on the active and standby OLTs are the same.

 Copyright © Huawei Technologies Co., Ltd.

Precaution

Users must manually configure the same service data on the two OLTs so that the ONU can rapidly switch services from the active OLT to the standby one, minimizing service interruption duration.

NOTE

Configurations of xPON network protection do not affect service configuration. This topic describes how to perform configurations of xPON network protection. For details about how to perform service configuration, see the specific service configuration guide.

Procedure

Step 1 (Optional)Shutdown the laser of the protection member PON port on the Secondary OLT (OLT-B).

NOTE

This step applies to the following scenarios, Otherwise, skip this step.

- The ONU upstream port works in auto-negotiation mode. That is, the ONU upstream port can work in GPON mode or XG(S)-PON mode according to the port type on the OLT side.
- The optical fiber of the protection member PON port of Type X protection is connected, that is, the physical link is connected. In this case, the physical link is connected but the Type X protection data configuration is not complete. The working and protection member PON ports transmit light at the same time and interact with the ONU. As a result, the upstream port mode of the ONU is repeatedly switched and the ONU is repeatedly reset.

To solve the preceding problem:

1. Run the **shutdown** command to disable the laser on the protection member PON port before the configuration.
2. After the configuration is complete, run the **undo shutdown** command to turn on the laser of the protection member PON port.

Step 2 Configure a dual-homing xPON type B protection group on the active OLT (OLT-A).

1. Run the **dual-parenting local-node** command to configure the local IP address, TCP port number, and key.
2. Run the **dual-parenting peer-node** command to configure the peer IP address, TCP port number, and key.
3. Run the **dual-parenting sync** command to enable dual-homing synchronization.
4. Run the **protect-group** command to create a protection group.

   Run the **description** command to configure the description of a PG.
5. Run the **protect-group member** command to add a member to the protection group.

   NOTE

   After a member is added to a dual-homing protection group, the group is automatically enabled.
6. Run the **peer-group-member** command to configure the peer member of the protection group.

Step 3 Configure a dual-homing xPON type B protection group on the standby OLT (OLT-B).

 Copyright © Huawei Technologies Co., Ltd.

1. Run the **dual-parenting local-node** command to configure the local IP address, TCP port number, and key.
2. Run the **dual-parenting peer-node** command to configure the peer IP address, TCP port number, and key.
3. Run the **dual-parenting sync** command to enable dual-homing synchronization.
4. Run the **protect-group** command to create a protection group.

   The description of the protection groups created on the active and standby OLTs must be the same. Run the **description** command to configure the description of a PG.
5. Run the **protect-group member** command to add a member to the protection group.
6. Run the **peer-group-member** command to configure the peer member of the protection group.
7. Run the **protect-group enable** command to enable the protection group.

Step 4 (Optional) Run the **uplink-monitor port** command to associate the protection groups with the uplink Ethernet port status.

NOTE

If an Ethernet link aggregation group has been configured, make sure that the Ethernet port associated with the dual-homing protection groups is the master port in the aggregation group.

- Run the **uplink-monitor bfd** command to associate the protection groups with a BFD session. For instructions about how to configure a BFD session, see **Configuring a BFD Session**.
- Run the **uplink-monitor mep** command to associate the protection groups with an MEP session. For instructions about how to configure an MEP session, see **Configuring CFM**.

NOTE

One protection group can be configured only one fault detection type.

Step 5 (Optional)Turn on the laser of the standby PON port. After the preceding configuration steps are complete, run the **undo shutdown** command to turn on the laser of the protection member PON port.

----End

Result

After the configuration, both active and standby OLTs are in working state, and they both check link status for determining a protection switchover.

An automatic switchover can be triggered by any of the following conditions:

- Optical fiber cut from the active OLT
- Active OLT PON port fault
- Active OLT fault
- Active line quality deterioration
- Active OLT uplink failure

 Copyright © Huawei Technologies Co., Ltd.

Example

The following section uses an example to describe how to configure a dual-homing GPON type B protection group on the active OLT (OLT_A) and standby OLT (OLT_B), respectively:

NOTE

The data in the example is for reference only. Replace it with the service data in the actual configuration.

- GPON service port on the two OLTs: 0/2/0
- Index of the protection groups on the two OLTs: 1
- Associated port with the dual-homing protection groups: 0/3/0
- IP address of the local OLT_A and peer OLT_B: 192.168.68.1; TCP port number: 6076; key: HW_42340000
- IP address of the local OLT_B and peer OLT_A: 192.168.68.8; TCP port number: 6076; key: HW_42340000

```
Configuration on OLT_B:
huawei_B(config)#interface gpon 0/2
huawei_B(config-if-gpon-0/2)#shutdown 0
huawei_B(config-if-gpon-0/2)#quit

Configuration on OLT_A:
huawei_A(config)#dual-parenting local-node ip-address 192.168.68.1 port 6076 key HW_42340000
huawei_A(config)#dual-parenting peer-node standby ip-address 192.168.68.8 port 6076 key
HW_42340000
huawei_A(config)#dual-parenting sync enable
huawei_A(config)#protect-group 1 protect-target gpon-uni-port workmode dual-parenting
huawei_A(protect-group-1)#protect-group member port 0/2/0 role work
huawei_A(protect-group-1)#peer-group-member peer-node standby peer-port 0/2/0
huawei_A(protect-group-1)#uplink-monitor port 0/3/0
huawei_A(protect-group-1)#quit
Configuration on OLT_B:
huawei_B(config)#dual-parenting local-node ip-address 192.168.68.8 port 6076 key HW_42340000
huawei_B(config)#dual-parenting peer-node active ip-address 192.168.68.1 port 6076 key
HW_42340000
huawei_B(config)#dual-parenting sync enable
huawei_B(config)#protect-group 1 protect-target gpon-uni-port workmode dual-parenting
huawei_B(protect-group-1)#protect-group member port 0/2/0 role protect
huawei_B(protect-group-1)#peer-group-member peer-node active peer-port 0/2/0
huawei_B(protect-group-1)#protect-group enable
huawei_B(protect-group-1)#uplink-monitor port 0/3/0
huawei_B(protect-group-1)#quit
huawei_B(config)#interface gpon 0/2
huawei_B(config-if-gpon-0/2)#undo shutdown 0
huawei_B(config-if-gpon-0/2)#quit
huawei_B(config)#save
```

16.2.7 xPON Type B Protection Reference Standards and Protocols

Standard No.	Description
ITU-T G.984.1	General Characteristics. This protocol mainly describes the basic features and major protection modes of GPON.

 Copyright © Huawei Technologies Co., Ltd.

Standard No.	Description
ITU-T G.984.2	Physical Media Dependent (PMD) Layer Specification. This protocol mainly describes the PMD layer parameters, including physical parameters (such as the transmit optical power, receiver sensitivity, and overload optical power) of optical transceivers, and also defines optical budget of different levels, for example, the most common Class B+.
ITU-T G.984.3	Transmission Convergence Layer Specification. This protocol mainly describes the TC layer specifications, including the upstream and downstream frame structures and GPON principle.
ITU-T G.984.4	ONT Management And Control Interface Specification. This protocol mainly describes the GPON management and maintenance protocols, such as OAM, PLOAM, and OMCI.
ITU-T G.984.5	Enhancement Band. This protocol mainly describes the GPON wavelength planning, including reserving bands for next-generation PON.
ITU-T G.984.7	Long Reach. This protocol mainly describes several long reach PON schemes for extending GPON transmission distance.
ITU-T G.988	ONU management and control interface (OMCI) specification.
TR-156	Using GPON Access in the context of TR-101.

16.3 xPON Type C Protection

The xPON type C protection switching is implemented through the redundancy configuration of the two PON ports on the ONU, backbone optical fiber, optical splitter, and tributary optical fiber on a xPON network. Each item is in a dual configuration. The protection improves the reliability on the optical distribution network (ODN) and prevents service interruption.

 Copyright © Huawei Technologies Co., Ltd.

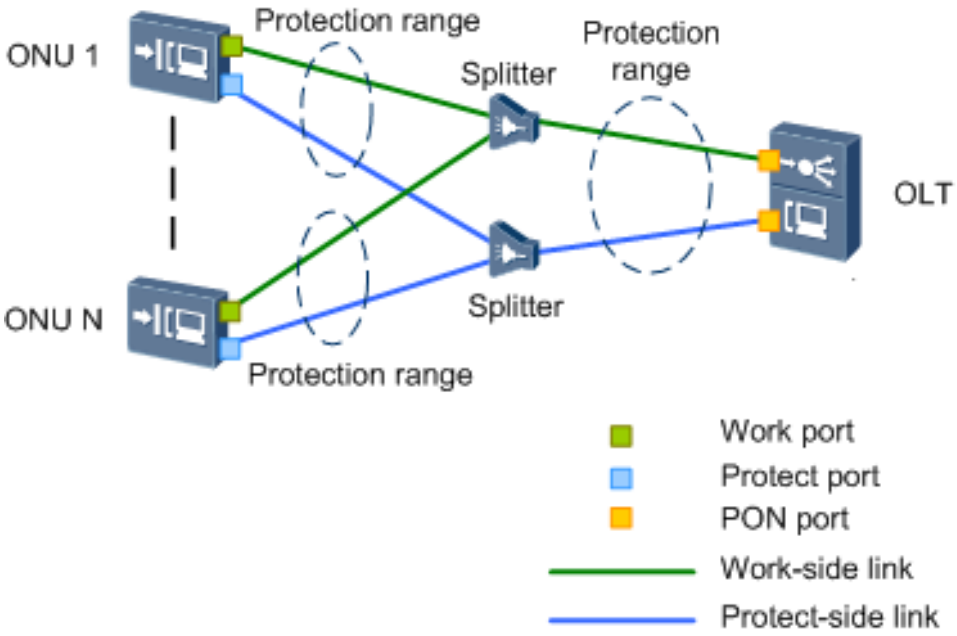
16.3.1 Introduction to xPON Type C Protection

Type C provides redundancy for OLT (dual homing), ONU's PON ports, backbone fibers, optical splitters, and distribution optical fibers. When a fault occurs, services can be automatically switched to the functional link. After the fault is rectified, services are automatically switched back to the original link.

xPON type C protection can be deployed in two networking scenarios: single homing and dual homing.

Figure 16-11 shows the xPON type C protection single homing network.

Figure 16-11 xPON type C protection network (single homing)

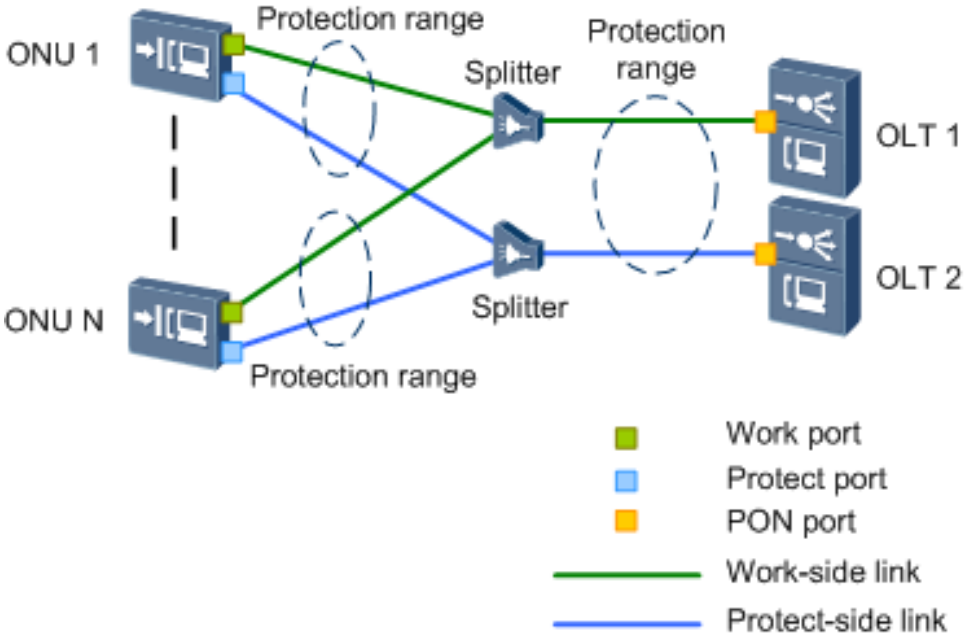


Networking Mode	Advantage	Disadvantage
Single homing	The networking mode is simple, and OLT and ONU can be managed easily.	When the OLT becomes faulty, services are interrupted. Optical fibers are deployed on the same channel and therefore two optical fibers may be broken at the same time.

Figure 16-12 shows the xPON type C protection dual homing network.

 Copyright © Huawei Technologies Co., Ltd.

Figure 16-12 xPON type C protection network (dual homing)



Networking Mode	Advantage	Disadvantage
Dual homing	When the active OLT or its uplink fails, services can be switched to the standby OLT.	The networking mode is complicated and costly, and the ONU management is difficult.

16.3.2 Basic Concepts of xPON Type C Protection

Protection Group

On a single-homed network, two PON uplink ports on an ONU connected to different PON ports on an OLT are added to a protection group.

NOTE

The OLT PON ports can be on the same board or on different boards. The differences are as follows:

- Port redundancy backup on the same board can conserve hardware resources. If the PON service board fails, the services on the entire board are interrupted.
- Port redundancy backup on the different boards requires hardware costs than that on the same board. If the active PON service board fails, the services can be automatically switched over to the PON ports on the standby board without being interrupted.

On a dual-homed network, two PON uplink ports on an ONU connected to two OLTs are added to a protection group. Switching can be performed between two members in a protection group.

 Copyright © Huawei Technologies Co., Ltd.

Roles of Protection Group Members

Protection group members have two roles: working and protection.

- One protection group contains a working port (the member's role is working) and a protection port (the member's role is protection). The working port and protection port are two different uplink PON ports on the ONU.
- In normal cases, the working port carries services.
- When the link of the working port becomes faulty, the system automatically switches services from the working port to the protection port to ensure service continuity.

State of Protection Group Members

Protection group members have two states: active and standby. The active port forwards data and the standby port does not forward data.

Switching Types

The switching can be triggered automatically by a fault or performed manually. Manual operations that may cause switching are manual switching, forcible switching, and locking.

- In automatic switching, the OLT and ONU automatically switch to the standby link when the conditions for triggering the switching are met.
- In manual switching, users manually switch the protection group by running the **manual-switch** command on the OLT.
- In forcible switching, users run the **force-switch** command on the OLT to perform the switching regardless of whether specific group members are running properly.
- After switching, the working port's state becomes standby. Then, if users run the **lockout** command on the OLT to lock a group member (only the protection port can be locked), the switching is performed and the working port's state becomes active.

NOTE

- In training switching, users run the **exercise-switch** command on the OLT to perform the switching to test the Automatic Protection Switching (APS) function on the ports in a protection group. Services are not switched.
- In automatic switchback, when the working member in the PG recovers to the normal state, the PG automatically switches over after the WTR time expires, and service is still carried on the working member.

Switching Condition

Protection group members are switched only when the following conditions are met:

- The protection group is enabled.

  NOTE

  The status of a protection group can be queried using the **display protect-group** command on the OLT. If **Admin State** is displayed in the output, the protection group is enabled.

 Copyright © Huawei Technologies Co., Ltd.

- The protection group is not locked using the **lockout** command on the OLT.
- The protection group member is not forcibly switched using the **force-switch** command on the OLT.

NOTE

Type C protection requires the cooperation of the OLT and ONU to implement active/standby switching. Third-party ONUs have not been tested with Huawei OLTs. Contact Huawei maintenance engineers to check whether they can work properly.

Operation Restriction Relationships in Protection Switching

Table 16-4 Type C single homing protection switching

Current Status			Next Operation									Remarks
Enabled	Locked	Forcible switching	Enabled	Disabled	Locked	Unlocked	Forcible switching	Canceling forcible switching	Automatic switching	Manual switching	Training switching	None
No	No	No	Supported	N/A	Supported	N/A	N/A	N/A	N/A	N/A	N/A	None
No	Yes	No	Supported	N/A	N/A	Supported	N/A	N/A	N/A	N/A	N/A	None
Yes	No	No	Not supported	Supported	Supported	N/A	Supported	N/A	Supported	Supported	Supported	None

Current Status			Next Operation									Remarks
Yes	No	Yes	Not supported	Supported	Supported	N/A	Supported	Supported	N/A	N/A	N/A	The forcible switching status will be cleared when the protection switching is disabled or locked.
Yes	Yes	No	Not supported	Supported	N/A	Supported	N/A	N/A	N/A	N/A	N/A	None

NOTE

The statuses that are not listed in the preceding table, such as disabled and forcible switching, are unavailable.

 Copyright © Huawei Technologies Co., Ltd.

Table 16-5 Type C dual homing protection switching

Current Status			Next Operation									Remarks
Enabled	Locked	Forcible switching	Enabled	Disabled	Locked	Unlocked	Forcible switching	Canceling forcible switching	Automatic switching	Manual switching	Training switching	None

Current Status			Next Operation									Remarks
No	No	No	Supported	N/A	Supported	N/A	N/A	N/A	N/A	N/A	N/A	Protection switching is consistently enabled on the working side, and the supported statuses are available only on the protection side.

Current Status			Next Operation									Remarks
No	Yes	No	Supported	N/A	N/A	Supported	N/A	N/A	N/A	N/A	N/A	Protection switching is consistently enabled on the working side, and the supported statuses are available only on the protection side.

 Copyright © Huawei Technologies Co., Ltd.

Current Status			Next Operation									Remarks
Yes	No	No	Not supported	Supported	Supported	N/A	Supported	N/A	Supported	Supported	Supported	Protection switching can be disabled or locked only on the protection side.

Current Status			Next Operation									Remarks
Yes	No	Yes	Not supported	Supported	Supported	N/A	Supported	Supported	N/A	N/A	N/A	Protection switching can be disabled or locked only on the protection side. In either of the statuses, the forcible switching status will be cleared.

Current Status			Next Operation									Remarks
Yes	Yes	No	Not supported	Supported	N/A	Supported	N/A	N/A	N/A	N/A	N/A	Protection switching can be locked only on the protection side.

NOTE

The statuses that are not listed in the preceding table, such as disabled and forcible switching, are unavailable.

Associated Protection Switching

Associated switching is implemented on a dual-homed network as follows: A protection group is associated on the OLT with the uplink Ethernet port status and BFD/MEP session status. In such a case, when the OLT's upper-layer network or the Layer 2 OLT physical link fails, the OLT determines a protection switchover, ensuring service continuity. Associated protection switching applies on the network enabled with automatic site information transmission.

16.3.3 Single-Homing xPON Type C Protection Principles

On a single homing network, one ONU is connected to two PON ports on an OLT, one working as the active port and the other as standby. The two ports on the OLT cannot forward packets at the same time. An automatic switching can be triggered by any of the following conditions:

- Loss of signal (LOS) occurs in the input direction.
- The ONU is offline.
- The OLT or ONU hardware is faulty.

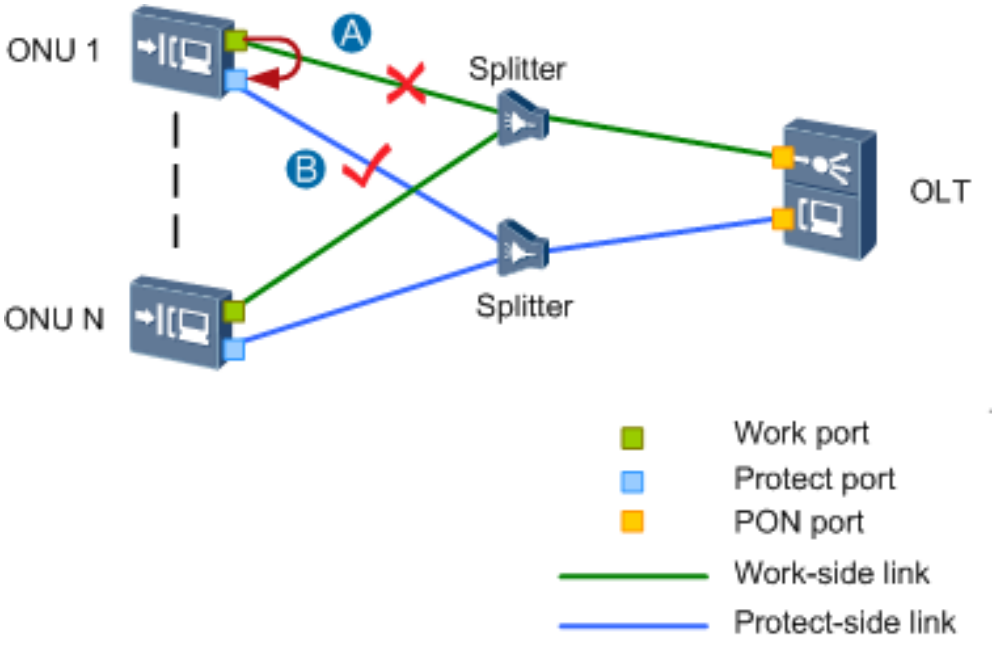
The following section describes PON switching processes in three scenarios. The OLT's PON ports are running properly, and the ONU has registered with the OLT. Users have issued the same settings of the PON line to PON port 1 and PON port 2 on the OLT so that services can recover after protection switching.

 Copyright © Huawei Technologies Co., Ltd.

Scenario 1: Branch Fiber Connected to a Single ONU Becomes Faulty

Figure 16-13 shows the scenario in which the branch fiber connected to a single ONU becomes faulty.

Figure 16-13 Branch fiber connected to a single ONU becomes faulty



1. Both the ONU and OLT check the link status and determine whether to trigger protection switching based on the link status.
   - If the OLT detects a fault on link A in Figure 16-13, it automatically switches to the protection link and uses the protection link to send messages to ONU 1 to notify that protection switching has occurred. In addition, the OLT notifies ONU 1 of the switching cause.
   - If ONU 1 detects a fault on link A in Figure 16-13, it automatically switches to the protection link and sends messages to the OLT to notify that protection switching has occurred. In addition, the ONU notifies the OLT of the switching cause.
2. After switching, services on ONU 1 are transmitted to the OLT through the protection port, all the feeder fibers connected to the OLT transmitted service packets, and ONU N is not affected. The changes on ONU 1 are as follows:
   - The state of the working port changes to standby.
   - The state of the protection port changes to active, and service packets are transmitted through link B in Figure 16-13.
3. After protection switching, ONU 1 can automatically switch back to the working port. The OLT sends an automatic switchback message and the switchback time, called the wait to restore (WTR) time, to the ONU.
   - If the OLT learns that ONU 1's working port and the working line are functioning properly and the working line stays normal during the WTR time, the OLT automatically switches back to the working line when the

 Copyright © Huawei Technologies Co., Ltd.

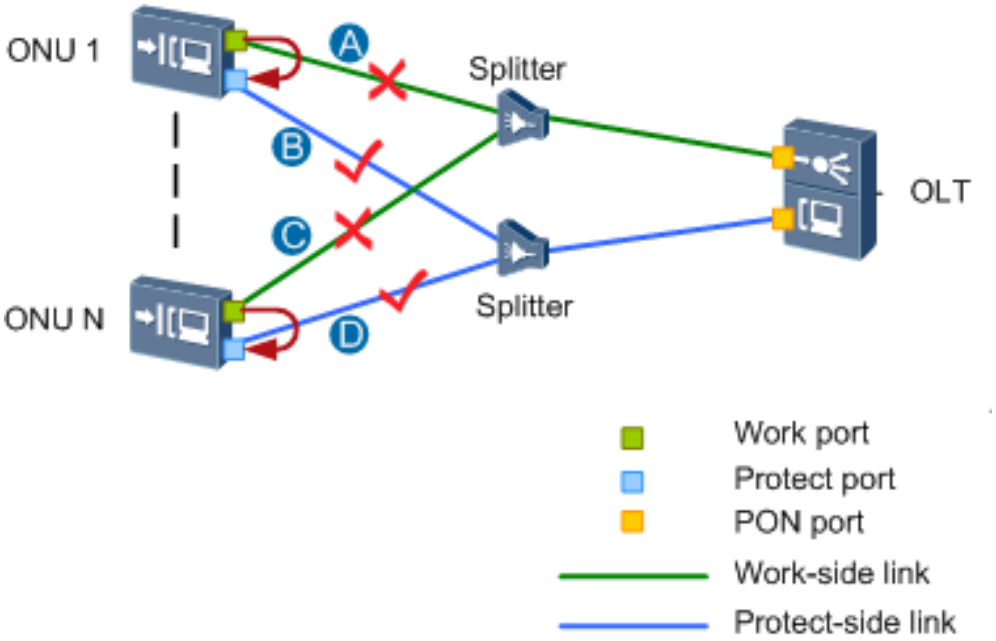
WTR time expires. In addition, the OLT notifies the ONU of the switching and switching cause.

- If ONU 1 learns that its working port and the working line are functioning properly and the working line stays normal during the WTR time, the ONU automatically switches back to the working line when the WTR time expires. In addition, the ONU notifies the OLT of the switching and switching cause.

Scenario 2: All Branch Fibers Connected to the ONU Become Faulty

Figure 16-14 shows the scenario in which all branch fibers connected to the ONU become faulty.

Figure 16-14 All branch fibers connected to the ONU become faulty



1. Both the ONU and OLT check the link status and determine whether to trigger protection switching based on the link status.
   - If the OLT detects that all branch fibers connected to the working line become faulty, it automatically switches to the protection link and uses the protection link to send messages to all the ONUs to notify that protection switching has occurred. In addition, the OLT notifies all the ONUs of the switching cause.
   - If the ONU detects a fault on all the branch fibers connected to the working link, it automatically switches to the protection link and sends messages to the OLT to notify that protection switching has occurred. In addition, the ONU notifies the OLT of the switching cause.
2. After switching, services on the ONU are transmitted to the OLT through the protection port (that is, service packets are transmitted by the protection link). The changes on the ONU are as follows:

 Copyright © Huawei Technologies Co., Ltd.

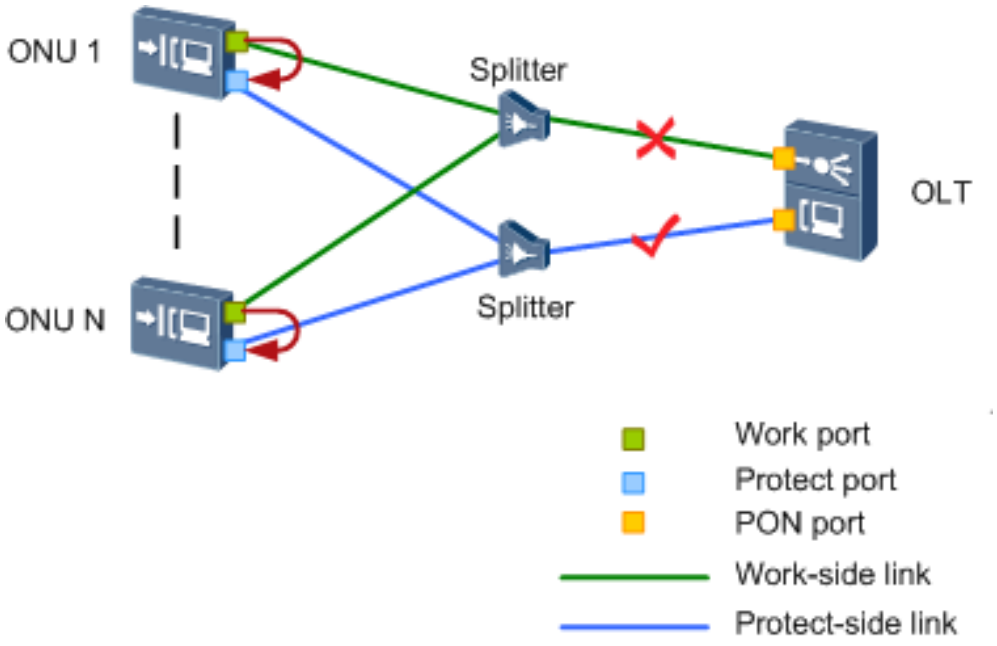
- The state of the working port changes to standby.
- The state of the protection port changes to active.

3. After protection switching, the ONU can automatically switch back to the working port. The OLT sends an automatic switchback message and the switchback time, called the WTR time, to the ONU.
   - If the OLT learns that the working port and working links are functioning properly and link A stays normal during the WTR time, the OLT automatically switches back to the working link when the WTR time expires. In addition, the OLT notifies the ONU of the switching and switching cause.
   - If the ONU learns that the working port and working link are functioning properly and link A stays normal during the WTR time, the ONU automatically switches back to the working link when the WTR time expires. In addition, the ONU notifies the OLT of the switching and switching cause.

Scenario 3: Backbone Fiber Becomes Faulty

Figure 16-15 shows the scenario in which the backbone fiber becomes faulty.

Figure 16-15 Backbone fiber becomes faulty



1. Both the ONU and OLT check the link status and determine whether to trigger protection switching based on the link status.
   - If the OLT detects a fault on the working link, it automatically switches to the protection link and uses the protection link to send messages to all the ONUs to notify that protection switching has occurred. In addition, the OLT notifies all the ONUs of the switching cause.

 Copyright © Huawei Technologies Co., Ltd.

- If the ONU detects a fault on the working link, it automatically switches to the protection link and sends messages to the OLT to notify that protection switching has occurred. In addition, the ONU notifies the OLT of the switching cause.

2. After switching, services on the ONU are transmitted to the OLT through the protection port (that is, service packets are transmitted by the protection link). The changes on the ONU are as follows:
   - The state of the working port changes to standby.
   - The state of the protection port changes to active.
3. After protection switching, the ONU can automatically switch back to the working port. The OLT sends an automatic switchback message and the switchback time, called the WTR time, to the ONU.
   - If the OLT learns that the working port and working links are functioning properly and link A stays normal during the WTR time, the OLT automatically switches back to the working link when the WTR time expires. In addition, the OLT notifies the ONU of the switching and switching cause.
   - If the ONU learns that the working port and working link are functioning properly and link A stays normal during the WTR time, the ONU automatically switches back to the working link when the WTR time expires. In addition, the ONU notifies the OLT of the switching and switching cause.

16.3.4 Dual-Homing xPON Type C Protection Principles

On a dual homing network, two PON lines, one working as the active line and one as standby, between an ONU and two OLTs cannot forward packets at the same time. An automatic switchover can be triggered by any of the following conditions:

- Loss of signal (LOS) occurs in the input direction.
- The ONU is offline.
- The OLT or ONU hardware is faulty.
- The OLT's uplink is faulty (this condition triggers automatic switching only in the associated protection switching scenario).

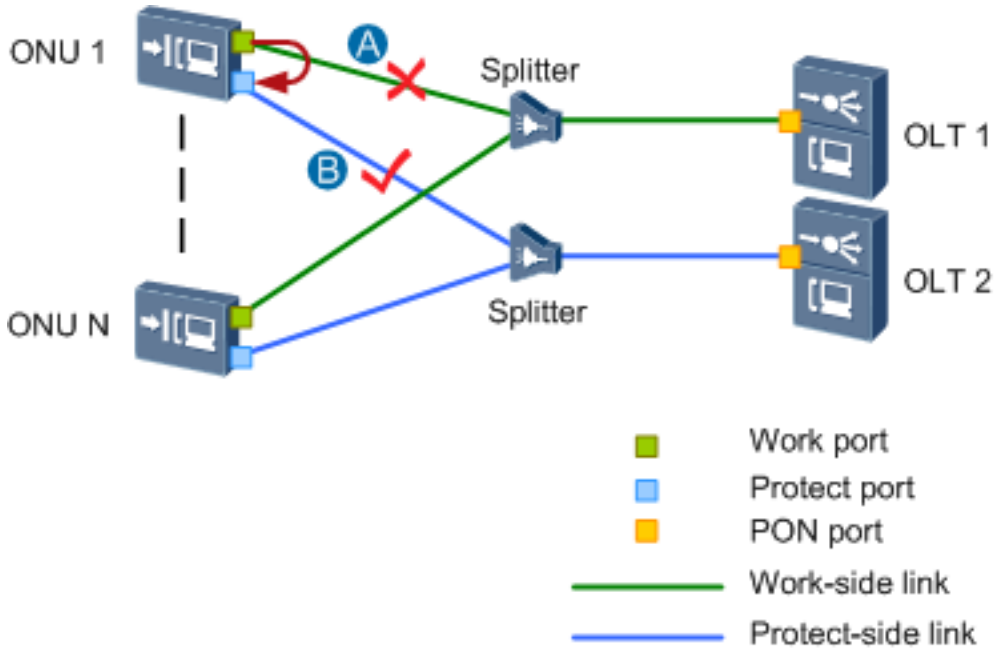
The following section describes the PON switching processes in five scenarios. OLT 1 and OLT 2 are running properly, and the ONU has registered with the OLTs. Users issue the PON's configurations to OLT 1 and OLT 2 so that ONU's services can recover after the switching.

Scenario 1: Branch Fiber Connected to a Single ONU Becomes Faulty

Figure 16-16 shows the scenario in which the branch fiber connected to a single ONU becomes faulty.

Figure 16-16 Branch fiber connected to a single ONU becomes faulty

 Copyright © Huawei Technologies Co., Ltd.



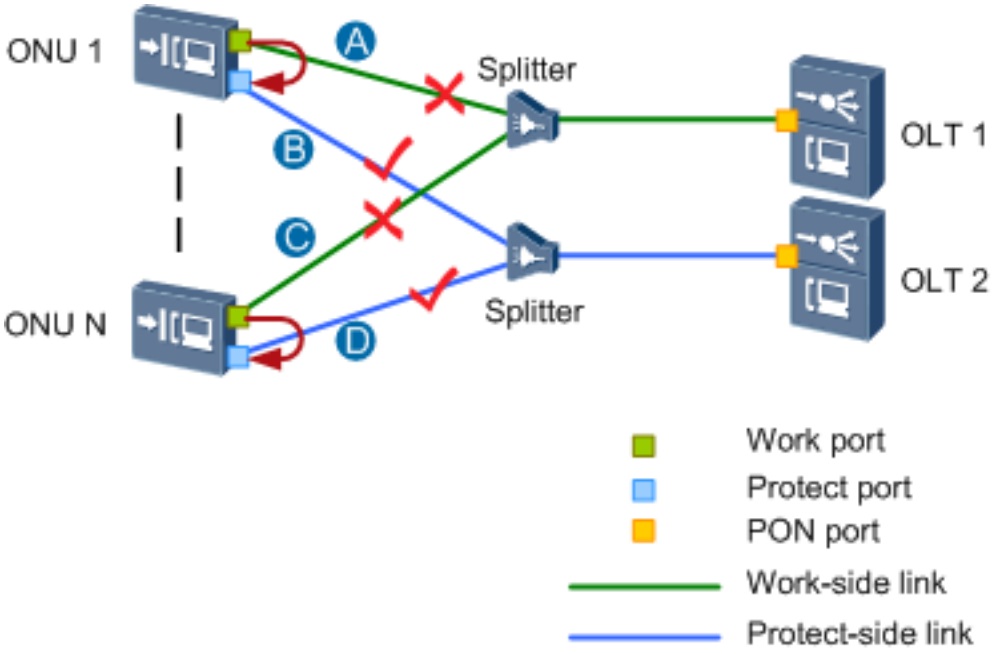
- If ONU 1 detects a fault on link A in **Figure 16-16**, it automatically switches to the protection link and sends messages to OLT 2 to notify that protection switching has occurred. In addition, the ONU notifies OLT 2 of the switching cause.
- After switching, services on ONU 1 are transmitted to the OLT through the protection port, all the backbone fibers connected to the OLT transmit service packets, and ONU N is not affected. The changes on ONU 1 are as follows:
  - The state of the working port changes to standby.
  - The state of the protection port changes to active and service packets are transmitted through link B in **Figure 16-16**.
- After protection switching, ONU 1 can automatically switch back to the working port. The OLT sends an automatic switchback message and the switchback time, called the wait to restore (WTR) time, to the ONU. If ONU 1 learns that the working port, working link, and the uplink of OLT 1 are functioning properly and link A stays normal during the WTR time, ONU 1 automatically switches to the working link when the WTR time expires.

Scenario 2: All Branch Fibers Connected to the Active Link Become Faulty

Figure 16-17 shows the scenario in which all branch fibers connected to the active link become faulty.

 Copyright © Huawei Technologies Co., Ltd.

Figure 16-17 All branch fibers connected to the active link become faulty



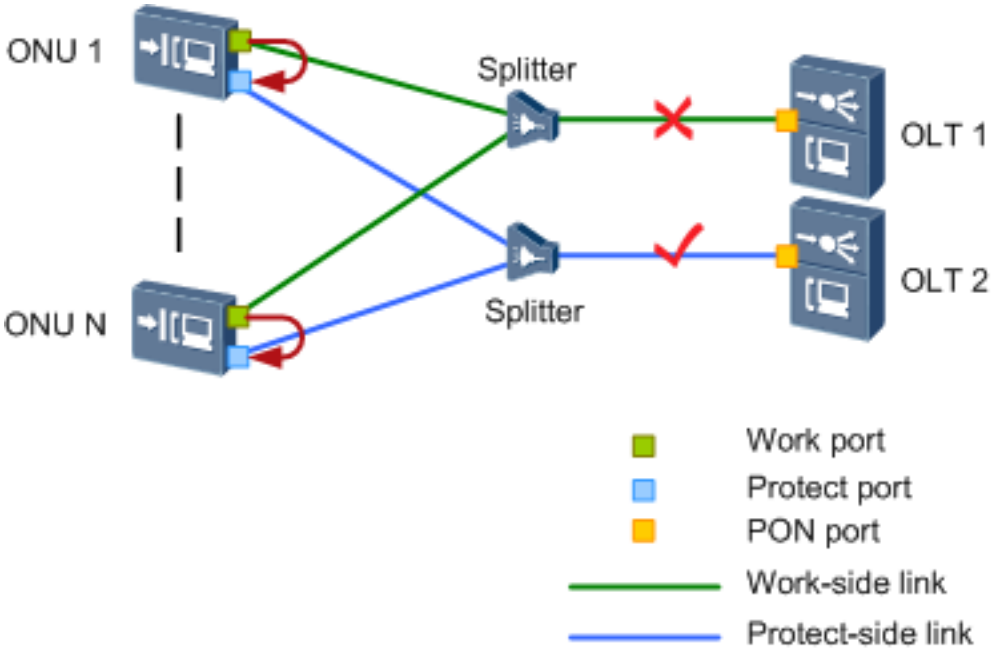
- If the ONU detects a fault on all the branch fibers connected to the working link, it automatically switches to the protection link and sends messages to OLT 2 to notify that protection switching has occurred. In addition, the ONU notifies OLT 2 of the switching cause.
- After switching, services on the ONU are transmitted to OLT 2 through the protection port (that is, service packets are transmitted by the protection link). The changes on the ONU are as follows:
  - The state of the working port changes to standby.
  - The state of the protection port changes to active.
- After protection switching, the ONU can automatically switch back to the working port. The OLT sends an automatic switchback message and the switchback time, called the WTR time, to the ONU. If the ONU learns that the working port, working links, and the uplink of OLT 1 are functioning properly and link A and link C stays normal during the WTR time, the ONU automatically switches to the working links when the WTR time expires.

Scenario 3: Backbone Fiber Becomes Faulty

Figure 16-18 shows the scenario in which the backbone fiber becomes faulty.

 Copyright © Huawei Technologies Co., Ltd.

Figure 16-18 Backbone fiber becomes faulty



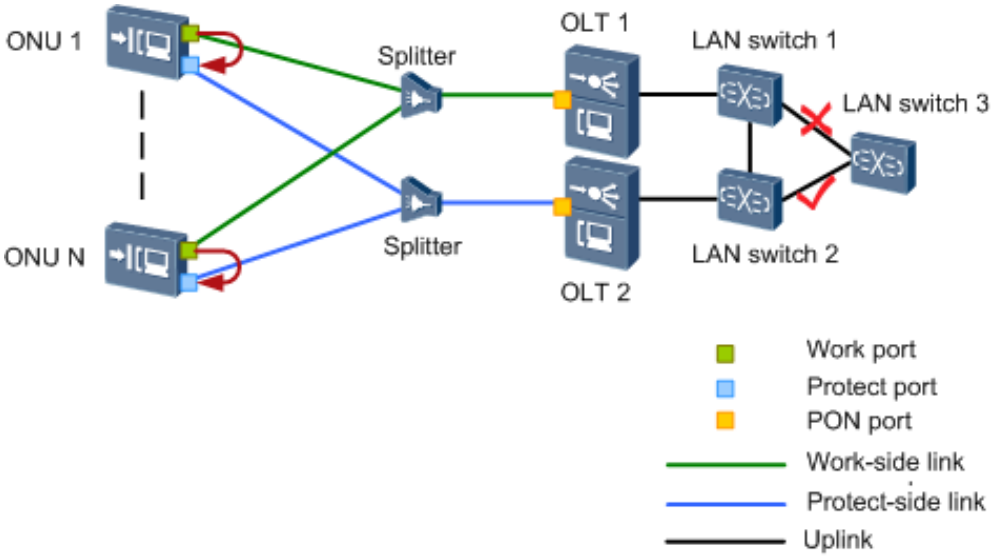
- If the ONU detects a fault on the backbone fiber, it automatically switches to the protection link and sends messages to OLT 2 to notify that protection switching has occurred. In addition, the ONU notifies OLT 2 of the switching cause.
- After switching, services on the ONU are transmitted to OLT 2 through the protection port (that is, service packets are transmitted by the protection link). The changes on the ONU are as follows:
  - The state of the working port changes to standby.
  - The state of the protection port changes to active.
- After protection switching, the ONU can automatically switch back to the working port. The OLT sends an automatic switchback message and the switchback time, called the WTR time, to the ONU. If the ONU learns that the working port, working links, and the uplink of OLT 1 are functioning properly and link A and link C stays normal during the WTR time, the ONU automatically switches to the working links when the WTR time expires.

Scenario 4: Associated Protection Switching Caused by a Fault on the OLT's Uplink

An OLT protection group is associated with the BFD or MEP session. Based on the association, when the upper-layer network connection (or IP layer link) of the OLT fails, the OLT instructs the ONU to trigger protection switching, which ensures service continuity. **Figure 16-19** shows the associated protection switching caused by a fault on the OLT's uplink.

 Copyright © Huawei Technologies Co., Ltd.

Figure 16-19 Associated protection switching caused by a fault on the OLT's uplink



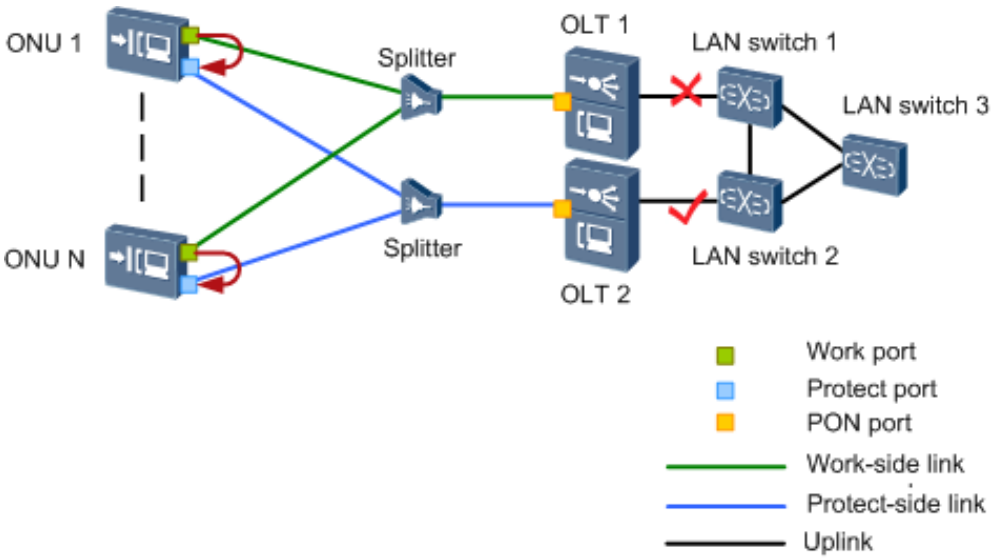
- On the OLT, the dual-homing protection group is associated with the BFD or MEP session. If the upstream route of the OLT fails, the BFD session status becomes **Down** or the MEP detection fails. In such a case, the OLT notifies the ONU of the uplink change.
- After the ONU receives the switching instruction from the OLT, it determines to trigger the switching and switches to the protection link. After the switching, the ONU notifies OLT 1 of the switching using the working link and notifies OLT 2 of the switching using the protection link.
- After switching, services on the ONU are transmitted to OLT 2 through the protection port (that is, service packets are transmitted by the protection link). The changes on the ONU are as follows:
  - The state of the working port changes to standby.
  - The state of the protection port changes to active.
- After protection switching, the ONU can automatically switch back to the working port. The OLT sends an automatic switchback message and the switchback time, called the WTR time, to the ONU. If the ONU learns that the working port, working links, and the uplink of OLT 1 are functioning properly and link A and link C stays normal during the WTR time, the ONU automatically switches to the working links when the WTR time expires.

Scenario 5: Associated Protection Switching Caused by a Fault on the OLT's Layer 2 Physical Link

An OLT protection group is associated with the uplink Ethernet port status. Based on the association, when the Layer 2 physical link of the OLT fails, the OLT instructs the ONU to trigger protection switching, which ensures normal service transmission. **Figure 16-20** shows the associated protection switching caused by a fault on the OLT's Layer 2 physical link.

 Copyright © Huawei Technologies Co., Ltd.

Figure 16-20 Associated protection switching caused by a fault on the OLT's Layer 2 physical link



- On the OLT, the protection group is associated with the uplink Ethernet port status. When the Ethernet port associated with the protection group becomes **Down**, the OLT notifies the ONU of the uplink change.
- After the ONU receives the switching instruction from the OLT, it determines to trigger the switching and switches to the protection link. After the switching, the ONU notifies OLT 1 of the switching using the working link and notifies OLT 2 of the switching using the protection link.
- After switching, services on the ONU are transmitted to OLT 2 through the protection port (that is, service packets are transmitted by the protection link). The changes on the ONU are as follows:
  - The state of the working port changes to standby.
  - The state of the protection port changes to active.
- After protection switching, the ONU can automatically switch back to the working port. The OLT sends an automatic switchback message and the switchback time, called the WTR time, to the ONU. If the ONU learns that the working port, working links, and the uplink of OLT 1 are functioning properly and link A and link C stays normal during the WTR time, the ONU automatically switches to the working links when the WTR time expires.

16.3.5 Configuring xPON Type C Single-Homing Protection

This section describes how to configure xPON type C single-homing protection. Each optical network unit (ONU) provides two xPON uplink ports and connects to two xPON ports on an OLT through different optical splitters. This protects feeder and drop fibers and ensures high network reliability.

Context

After xPON type C single-homing protection is configured, the service configurations on the ONUs remain unchanged and data is transmitted or

 Copyright © Huawei Technologies Co., Ltd.

received over the active uplink ports on the ONUs and the active xPON port on the OLT.

NOTE

Configurations of xPON network protection do not affect service configuration. This topic describes how to perform configurations of xPON network protection. For details about how to perform service configuration, see the specific service configuration guide.

Precaution

Procedure

Step 1 Run the **ont add** command to add a working ONU.

Step 2 Run the **ont add** *portid ontid* **protect-side** command to add a protection ONU.

Ensure that the **protect-side** parameter is selected.

Step 3 Run the **protect-group protect-target gpon-uni-ont** command to add a xPON Type C protection group. The working mode of the protection group can only be **portstate**.

Step 4 Run the **protect-group member** command to add the working ONU to the protection group as a working member.

Step 5 Run the **protect-group member** command to add the protection ONU to the protection group as a protection member.

NOTE

Ensure that **ont** *ontid* value is the ONT ID specified in **Step 1**.

Step 6 Run the **protect-group enable** command to enable the xPON protection group.

A created protection group is disabled by default.

----End

Result

After the configuration, both the active and standby OLT xPON ports work in active mode. Both the ONU and OLT check the link status and determine whether to perform switchover accordingly.

Automatic protection switchover is triggered if one of the following requirements is met:

- Loss of signal (LOS) occurs in the input direction.
- The OLT or ONU hardware is faulty.

Example

The following configurations are used as an example to configure GPON type C single-homing protection on an OLT and ONUs:

- Ports on the same GPON service board: 0/1/0 and 0/1/1
- Working link: connected to port 0/1/0

 Copyright © Huawei Technologies Co., Ltd.

- Protection link: connected to port 0/1/1
- ONU ID: 0
- ONU authentication mode: SN; SN: hwhw-10101500; management mode: OMCI
- ID of the line profile bound to the ONU: 10

```
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#ont add 0 0 sn-auth hwhw-10101500 omci ont-lineprofile-id 10
huawei(config-if-gpon-0/1)#ont add 1 0 protect-side
huawei(config-if-gpon-0/1)#quit
huawei(config)#protect-group protect-target gpon-uni-ont workmode portstate
huawei(protect-group-1)#protect-group member port 0/1/0 ont 0 role work
huawei(protect-group-1)#protect-group member port 0/1/1 ont 0 role protect
huawei(protect-group-1)#protect-group enable
huawei(protect-group-1)#quit
```

16.3.6 Configuring xPON Type C Dual-Homing Protection

Type C dual-homing protection protects any node between an optical network unit (ONU) and the two dual-homed OLTs.

Context

Precaution

Compared with xPON type C single-homing protection, xPON type C dual-homing protection features enhanced protection capabilities but more complicated networking and higher deployment costs.

NOTE

Configurations of xPON network protection do not affect service configuration. This topic describes how to perform configurations of xPON network protection. For details about how to perform service configuration, see the specific service configuration guide.

Procedure

Step 1 Configure xPON type C dual-homing protection on the active OLT.

1. Run the **ont add** command to add a working ONU.
2. Run the protect-group command to create a protection group.

   Run the **description** command to configure the description of a PG.
3. Run the **protect-group member** command to add a working member to the protection group.

   NOTE

   After this step is performed, the protection group is automatically enabled.

Step 2 Configure xPON type C dual-homing protection on the standby OLT.

1. Run the **ont add** command to add a protection ONU.

   NOTE

   Ensure that all profiles on the ONUs connected to the active and standby OLTs are the same. The profiles include the dynamic bandwidth allocation (DBA) profile and line profile.

 Copyright © Huawei Technologies Co., Ltd.

2. Run the **protect-group** command to create a protection group.

   Run the **description** command to configure the description of a PG.
3. Run the **protect-group member** command to add a protection member to the protection group.
4. Run the **protect-group enable** command to enable the protection group.

----End

Result

After the configuration, both the active and standby OLTs are in the working state. The ONU and active/standby OLTs check the link status to determine whether to perform switchover accordingly.

An automatic switchover can be triggered by any of the following conditions:

- Loss of signal (LOS) occurs in the input direction.
- The OLT or ONU hardware is faulty.

Example

The following configurations are used as an example to configure GPON type C dual-homing protection on two OLTs:

- Active OLT: OLT_A; standby OLT: OLT_B
- GPON service ports: 0/1/0 on the two OLTs
- ID of the protection groups: 1
- ONU ID: 0
- ONU authentication mode: SN; SN: hwhw-10101500; management mode: OMCI
- ID of the line profile bound to the ONU: 10

```
OLT_A configurations:
huawei_A(config)#interface gpon 0/1
huawei_A(config-if-gpon-0/1)#ont add 1 0 sn-auth hwhw-10101500 omci ont-lineprofile-id 10
huawei_A(config-if-gpon-0/1)#quit
huawei_A(config)#protect-group 1 protect-target gpon-uni-ont workmode dual-parenting
huawei_A(protect-group-1)#protect-group member port 0/1/0 ont 0 role work
OLT_B configurations:
huawei_B(config)#interface gpon 0/1
huawei_B(config-if-gpon-0/1)#ont add 1 0 sn-auth hwhw-10101500 omci ont-lineprofile-id 10
huawei_B(config-if-gpon-0/1)#quit
huawei_B(config)#protect-group 1 protect-target gpon-uni-ont workmode dual-parenting
huawei_B(protect-group-1)#protect-group member port 0/1/0 ont 0 role protect
huawei_B(protect-group-1)#protect-group enable
```

16.3.7 xPON Type C Protection Reference Standards and Protocols

Standard No.	Description
ITU-T G.984.1	General Characteristics. This protocol mainly describes the basic features and major protection modes of GPON.

 Copyright © Huawei Technologies Co., Ltd.

Standard No.	Description
ITU-T G.984.2	Physical Media Dependent (PMD) Layer Specification. This protocol mainly describes the PMD layer parameters, including physical parameters (such as the transmit optical power, receiver sensitivity, and overload optical power) of optical transceivers, and also defines optical budget of different levels, for example, the most common Class B+.
ITU-T G.984.3	Transmission Convergence Layer Specification. This protocol mainly describes the TC layer specifications, including the upstream and downstream frame structures and GPON principle.
ITU-T G.984.4	ONT Management And Control Interface Specification. This protocol mainly describes the GPON management and maintenance protocols, such as OAM, PLOAM, and OMCI.
ITU-T G.984.5	Enhancement Band. This protocol mainly describes the GPON wavelength planning, including reserving bands for next-generation PON.
ITU-T G.984.7	Long Reach. This protocol mainly describes several long reach PON schemes for extending GPON transmission distance.
ITU-T G.988	ONU management and control interface (OMCI) specification.
TR-156	Using GPON Access in the context of TR-101.

16.4 Redundancy Backup of Control Boards

An OLT configured with two control boards works in the load sharing mode. When working normally, the active and standby control boards concurrently forward service packets.

 Copyright © Huawei Technologies Co., Ltd.

16.4.1 Introduction to Control Board Redundancy Backup

Definition

In this feature, two control boards, one working as the active and the other working as the standby, are configured to back up each other. When the two control boards work normally in load sharing mode, active and standby control boards concurrently forward service packets, and redundancy backup improves reliability of services as well as doubling bandwidth and enhancing data forwarding performance.

The redundancy backup of control boards benefits customers with the following capabilities:

- Quick switching between active and standby control boards
- Uninterrupted data forwarding during active/standby switchover
- Concurrent upgrade of the active and standby control boards

NOTE

The MA5800 configured with two control boards does not support the active/standby mode for the forwarding plane.

NOTE

In the case of dual control boards, the dual-uplink networking must be configured.

Introduction

The MA5800 configured with two control boards uses active/standby mode for the control plane and load sharing mode for the forwarding plane.

- On the control plane, the two control boards work in the active/standby mode. The CPU on the active control board manages the system and processes control protocol packets while the CPU on the standby control board is in the standby state.
- On the forwarding plane, the active and standby control boards share loads. Both boards forward data.

The MA5800 supports decoupling of the control plane and forwarding plane for the control board. When the **system switch-over** command is executed to trigger an active/standby switchover of the control plane, the forwarding plane of the standby control board continuously works and service bandwidths are not affected. This feature reduces impacts on services of the forwarding plane brought by a control plane reset in fault and upgrade scenarios for the control board, improving system In-Service Software Upgrade (ISSU) capabilities and reliability.

The MA5800 supports hot reset of the active or standby control board. During a hot reset, the forwarding plane continuously works and service bandwidths are not affected.

- Run the **reboot active hot** command to perform a hot reset on the active control board. After the CPU of the active control board resets, the system automatically switches to the standby control board. Then, the original active control board becomes the standby control board, ongoing online services are not offline, and bandwidths of existing services are not lowered.

 Copyright © Huawei Technologies Co., Ltd.

- Run the **reboot standby hot** command to perform a hot reset on the standby control board. After the CPU of the standby control board resets, ongoing online services are not offline, and bandwidths of existing services are not lowered.

16.4.2 Principle of Active/Standby Mode

Switchover Modes

- **Automatic active/standby switchover**

  When the active control board fails, the system automatically performs an active/standby switchover. Specifically, the system resets the active control board, and the standby control board functions as the new active control board. During the whole process, the system continues processing data and services are not interrupted.

- **Manual active/standby switchover**

  When you need to replace the active control board or upgrade the system software, manually perform an active/standby switchover using one of the following methods:

  – Run the **system switch-over** command.
  – Run the **reboot active [hot]** command to reset the active control board.
  – Press the RESET button on the active control board to reset the board.
  – Remove the active control board.

NOTICE

- During system running, if an active/standby switchover is performed by removing the active control board or pressing the RESET button, the hardware may be faulty or software data is not synchronized, leading to service failures. In this case, it is not recommended to remove the active control board or press the RESET button to perform an active/standby switchover. Specially, do not remove or install the active control board frequently in a short period of time.
- If the active control board has a daughter board and the standby control board does not have one, the BFD configuration will be lost after an active/standby switchover. Therefore, you are not advised to perform an active/standby switchover.

Communication Between the Active and Standby Control Boards

When two control boards are configured, the active control board is working and the standby control board is in the standby state. The standby control board exchanges information only with the active control board. The maintenance Ethernet port (designated with the ETH silkscreen) on the standby control board does not receive configuration commands. The serial port (designated with the CON silkscreen) on the standby control board receives configuration commands and forwards the commands to the active control board. Then the active control board issues the configurations to service boards. You can log in to the device only at the local end through the local serial port. The user name and password are

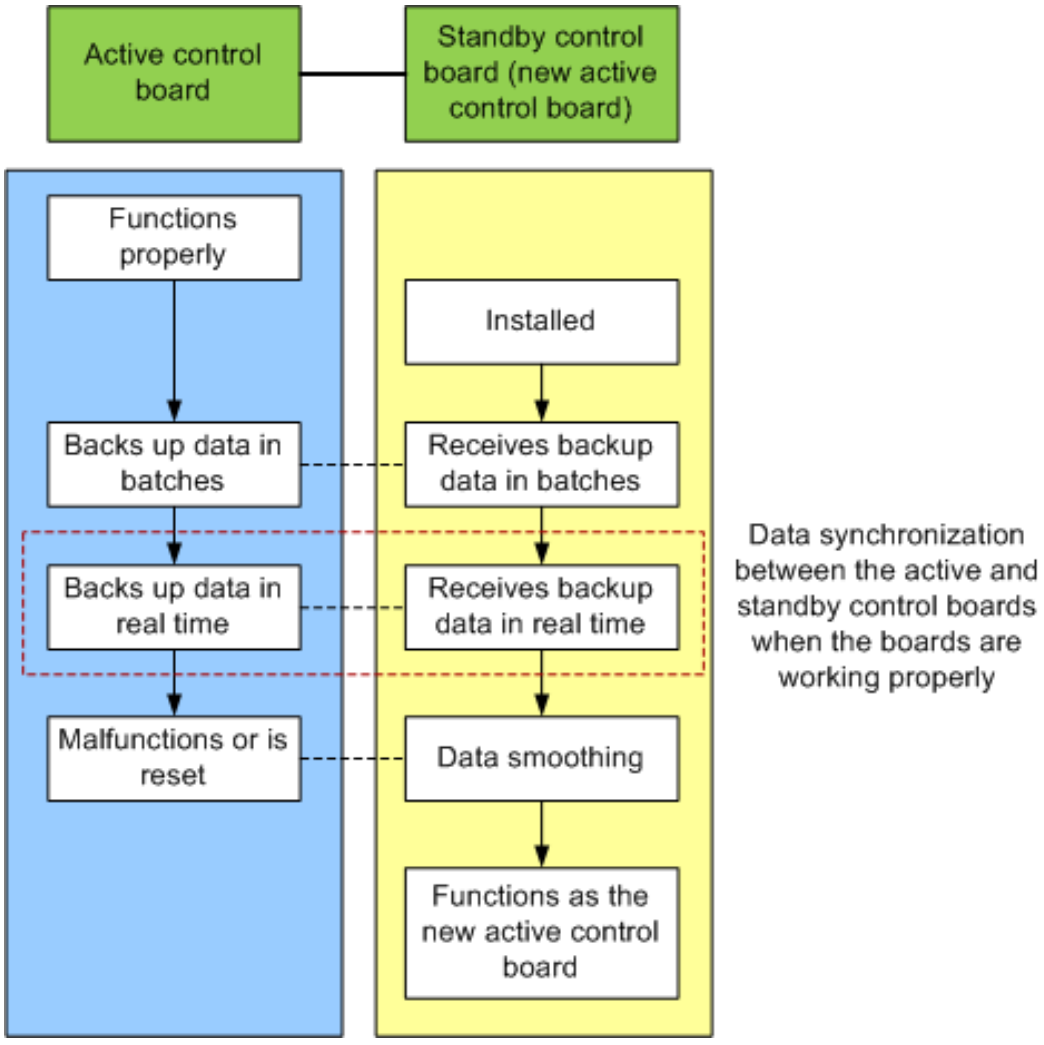
 Copyright © Huawei Technologies Co., Ltd.

used for authentication. The serial port is enabled by default and is usually used during device deployment. If the serial port is enabled for a long time, security risks may exist. You are advised to run the **sysman console disable** command to disable the serial port if it is not needed.

The active and standby control boards achieve data consistency between them using high availability (HA) mechanisms such as batch backup and real-time backup. When learning that the active control board fails or receiving an active/standby switchover command, the standby control board takes over as the active control board and reports switchover events.

Figure 16-21 shows the state transition between the active and standby control boards in active/standby mode.

Figure 16-21 State transition between the active and standby control boards



- **Batch backup**

  After the standby control board is installed, the active control board automatically backs up data of all modules to the standby control board in batches.

- **Real-time backup**

 Copyright © Huawei Technologies Co., Ltd.

After batch backup is complete, the system enters the real-time backup state. In this state, the active control board backs up data to the standby control board in real time. When both control boards function properly, the system stays in the real-time backup state.

Real-time backup maintains data consistency (also called data synchronization) between the active and standby control boards. The following data needs to be synchronized:

- Configuration data: core data, also the static data, running in the system, including all configurations manually issued by users and data that is generated based on user configurations in system initialization and ensures normal system running.
- Basic operating data: data that may change rarely during system running, including the device status (such as board status and port status), operation logs, and alarms. The basic operating data changes when the system status changes, for example, a board is faulty or a port connection status changes.
- Dynamic service data: data that changes in real time during system running, including data generated by a call service (such as a PPPoE dialup) and data changing quickly (such as an ARP entry change). The dynamic service data takes a majority of data to be synchronized, and it is the most difficult to be completely synchronized. Data smoothing is mainly performed for such data.

You can run the **display data sync state** command to query the data synchronization status (complete synchronization or incomplete synchronization).

- When all data on the standby control board is completely synchronized with that on the active control board, it is called complete synchronization.

  An active/standby switchover performed when data is completely synchronized is called a **normal switchover**. A normal switchover does not interrupt services. A normal switchover can be performed only after data (including configuration data, basic operating data, and dynamic service data) is completely synchronized and the cyclic redundancy check (CRC) results on the active and standby control boards are the same.
- When data on the standby control board is not completely synchronized with that on the active control board, it is called incomplete synchronization.

  An active/standby switchover performed when the data is not completely synchronized is called a **forced switchover**. **Table 16-6** lists three conditions of incomplete synchronization and whether forced switchover is supported in each condition.

 Copyright © Huawei Technologies Co., Ltd.

Table 16-6 Active/standby switchover in the case of incomplete synchronization

Data Incompletely Synchronized	Supports Forced Switchover by Commands	Supports Forced Switchover by Board Reset (Manually or Using Commands) or Board Removal	Switchover Results
Configuration data	No	Yes	The system resets and services are interrupted. Forced switchover is not recommended.
Basic data	No	Yes	The system does not reset but the service board may reset and some services are interrupted. Forced switchover is not recommended.
Some dynamic service data	Yes	Yes	Data that is not synchronized to the standby control board is lost after the switchover, but services are not affected, and data about connections, alarms, and logs is not lost.

- **Data smoothing**

  In the real-time backup state, if an active/standby switchover occurs, the standby control board will be promoted to the active role. Before the standby control board becomes active, modules on the standby control board collect and synchronize data from service boards. The data collection and synchronization process is called data smoothing. During data smoothing, modules on the standby control board actively communicate with service boards to confirm and synchronize hardware status, link layer status, and configuration data. In this manner, data and status information are consistent through the entire system so that the system can run normally after the switchover.

  The data smoothing duration is very short. After data smoothing is complete, the standby control board works as the new active control board.

Communication Between Service Boards and Control Boards

Service boards are connected to the active and standby control boards through buses. Data of service boards and data of the standby control board come from the active control board.

 Copyright © Huawei Technologies Co., Ltd.

When working normally, service boards and the active control board exchange control protocol packets. When an active/standby switchover occurs, service boards check out the switchover through hardware signals and then exchange control protocol packets with the original standby control board (now functions as the active control board).

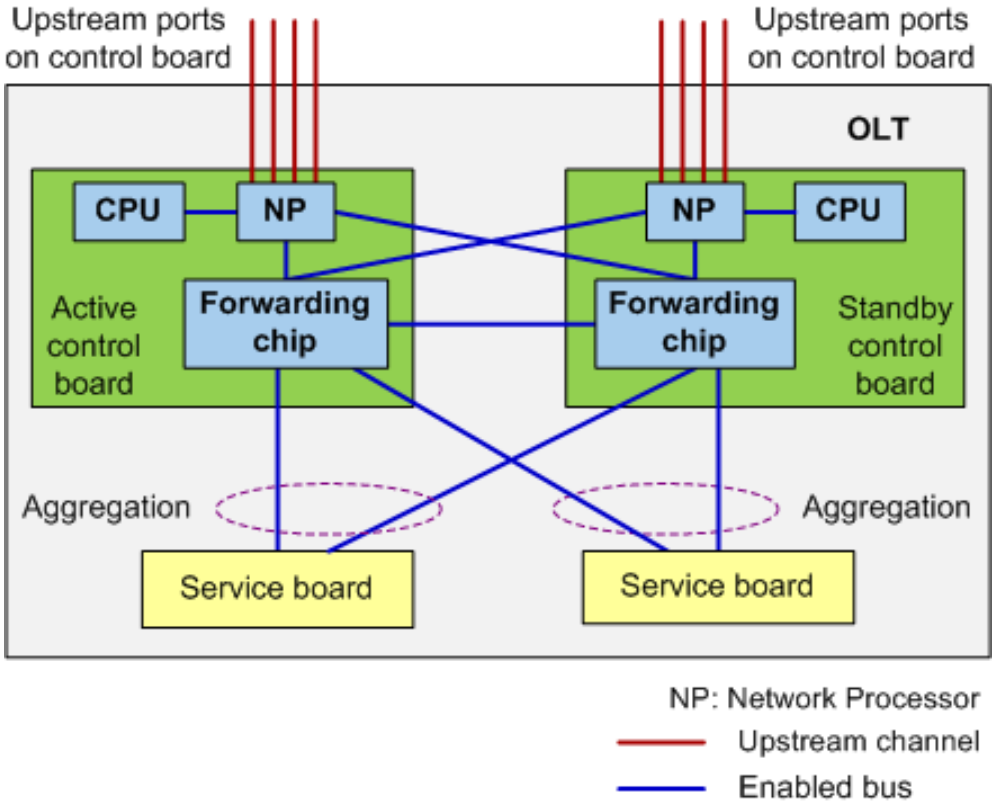
16.4.3 Principle of Load-sharing Mode

The MA5800 supports separation between the control plane and forwarding plane. When both active and standby control boards work normally, their CPUs manage their forwarding planes respectively. When the CPU of the active board is faulty, an active/standby switchover occurs, and the forwarding plane is not affected. When the CPU of the standby control board is faulty, the standby control board automatically restarts, and the forwarding plane is not affected either. The two control boards work in load sharing mode on forwarding planes, that is, both of them forward data and upstream ports are in working state at the same time. Also, traffic is sent from service boards to active and standby control boards in load sharing mode.

NOTE

When the main control board fails to start, both the control plane and forwarding plane fail.

Figure 16-22 Principle of load-sharing mode



 Copyright © Huawei Technologies Co., Ltd.

16.5 Ethernet Port Protection Group

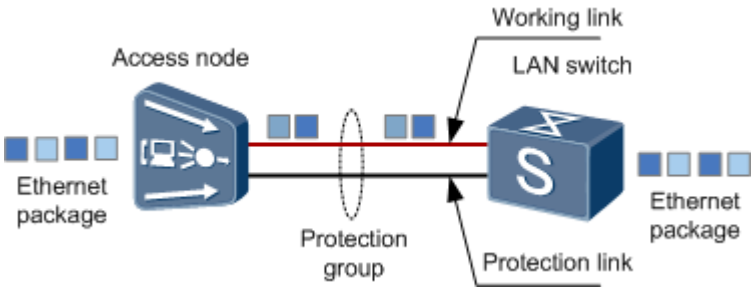
An Ethernet port protection group binds two or more Ethernet ports together. The working port carries services, while the protection port provides backup and does not carry services. If the working port is faulty, the services carried by it are switched to the protection port, automatically or manually as required. Ethernet port protection ensures uninterrupted service transmission and improves reliability of links.

16.5.1 Introduction to Protection Group of Ethernet Ports

Access network users, especially corporations and large residential communities, are posing stringent requirements on network reliability. If the upstream links of access devices are not protected, services carried over these upstream links will be interrupted if the links break or the upper-layer devices are malfunctioning. For the access devices that provide services to tens of thousands of users, carriers usually demand protection for the upstream ports. Specifically, services can be switched between links automatically or manually, depending on your requirements, for example, when the upstream link is broken. Such a protection group ensures that an upstream link is always available, improving reliability.

Figure 16-23 shows an example of an Ethernet port protection group between two interconnected devices. In this protection group, two Ethernet ports are bound. One Ethernet port protection group has two members. One member carries services over the working link, and the other member provides backup over the protection link.

Figure 16-23 Example of Ethernet port protection



According to the number of upstream Ethernet ports, members in a protection group can be ports or LAGs. Access devices support Timedelay Ethernet port protection group. A Timedelay Ethernet port protection group also switches services based on port status, but the protection port is disabled when the working port is functioning properly. If the working port is faulty, the protection port remains enabled for a predefined period of time. Services are switched to the protection port if this port is functioning properly during this period. If the protection port is malfunctioning during this period, the Timedelay protection group retries the working port while it keeps the protection port disabled. The protection group will repeat the preceding process until one port restores the normal state, and then switch services to it.

 Copyright © Huawei Technologies Co., Ltd.

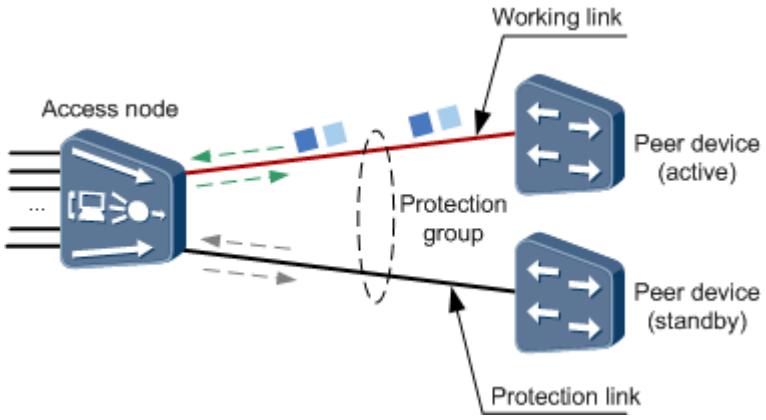
16.5.2 Principle of Timedelay Protection

Establishment of a Timedelay Protection Group

A Timedelay protection group is established after it is configured, and the interconnected devices implement protection without the need of negotiation.

In the protection group, the working port is enabled to transceive data and a service is transmitted over the link carried by the port; the protection port is disabled and no service is transmitted over the link carried by it. In the Timedelay protection scenario, the access device can be either single-homed or dual-homed to the upper-layer devices. **Figure 16-24** shows how data is transceived in the dual-homing scenario.

Figure 16-24 Data transceiving within a Timedelay protection group



Protection Switching

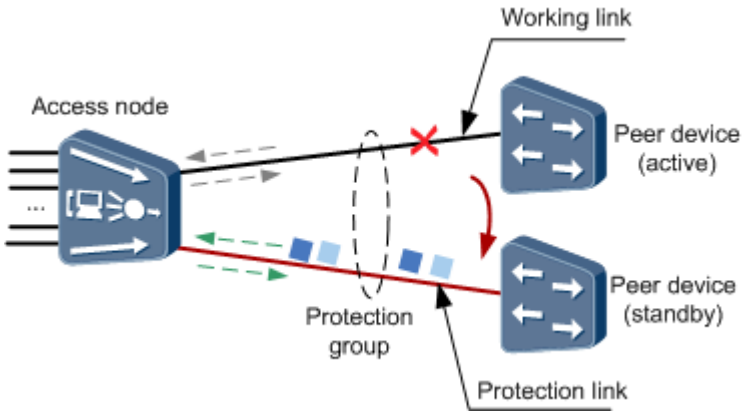
A Timedelay protection group undergoes protection switching on any of the following conditions:

- The working port is faulty.
- Users manually start a forced protection switching.

If the working link is faulty, services are switched to the protection link to ensure uninterrupted forwarding, as shown in **Figure 16-25**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 16-25 Data transceiving within a Timedelay protection group (after protection switching)



Timedelay protection switching is implemented as follows:

1. The access device detects a working port failure and immediately enables its protection port.
2. In the Timedelay protection group, the access device plays an "active" role and immediately switches services after detecting a working port failure. Services are then switched to the protection link.

NOTICE

The working and protection ports on peer devices must be enabled.

Protection Reversion

If the protection switching is triggered by a forced switching command or a working port failure, services will not be switched back to the original link even if the original working port restores the normal state. Instead, the services are still carried over the protection link.

The **reversion enable** command is used to enable the reversion function of a protection group.

- The priority for a working member in a protection group to carry the service is higher than the priority for a protection member in the protection group to carry the service. When a fault that occurs on the working member causes the protection switch, if the time for the fault of the working member to recover exceeds the WTR time, the switch reversion is performed. The protection switch refers the switch under which the working member switches to the protection member, and the switch reversion refers to the switch under which the protection member switches to the working member.
- If the switch of a protection group is initiated by a manual switch or a forced switch, the switch of the protection group cannot be reversed.
- If the cancel manual switch and cancel forced switch operations are performed and the working member is normal, the switch reversion operation is performed after the WTR time expires.

 Copyright © Huawei Technologies Co., Ltd.

16.5.3 Configuring a Timedelay Protection Group

A Timedelay protection group applies to the following scenario: The active and standby control boards or upstream service boards on an access device each provide an upstream Ethernet port, and users want port to carry services and the other port to back up. If the working port is faulty, the system switches services to the protection port to implement uninterrupted forwarding.

Prerequisites

- The boards and ports on the access device support Ethernet port protection groups.
- In the protection group, a static MAC address is configured only on the working port. You can run the **display mac-address** command to query the static MAC address.
- In a protection group, the two interconnected ports have the same data configurations and allow MAC address transfer.

Procedure

Step 1 (Optional) Configure the optical port shutdown function.

Run the **offline-tx-off-time** command to specify the time for keeping an optical port shut down in the case of a Linkdown. The optical port shutdown function helps improve protection switching performance.

Step 2 Create a Timedelay protection group.

Run the **protect-group** command to create a protection group (select **eth-nni-port** as the protection level and **Timedelay** as the protection type).

Step 3 Add a working port and a protection port to the protection group.

Run the **protect-group member** command to add one port on the control board or on the upstream board as the working port, and the other port on the board as the protection port in the protection group.

Step 4 Enable the protection group.

Run the **protect-group enable** command to enable the protection group.

Step 5 Query configurations about the protection group.

Run the **display protect-group** command to query configurations about the protection group. The configurations include the protection level, member ports, protection type, and reversion mode.

----End

Example

This example assumes a scenario in which the MA5800 connects to the upstream network through ports 0/9/0 and 0/9/1 which form a Timedelay protection group.

```
huawei(config)#protect-group 0 protect-target eth-nni-port workmode timedelay
huawei(protect-group-0)#protect-group member port 0/9/0 role work
huawei(protect-group-0)#protect-group member port 0/9/1 role protect
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(protect-group-0)#protect-group enable
huawei(protect-group-0)#display protect-group 0
  ------------------------------------------------------------------------
  Group ID         : 0
  Protect Target   : Port of Ethernet nni
  Work Mode        : timedelay
  Description      :
  Admin State      : enable
  Operation        : none
  Reversion        : disable
  Reversion Time(s): 720
  ------------------------------------------------------------------------
  Member       Role        Operation        State        PeerMember
  ------------------------------------------------------------------------
  0/9/0         work         none             active       none
  0/9/1       protect      none             standby      none
  ------------------------------------------------------------------------
```

16.6 Ethernet Link Aggregation

Ethernet link aggregation is a process that aggregates two or more Ethernet ports of the same type to a logic port. It increases the link bandwidth without requiring a hardware upgrade and improves link reliability using the link backup mechanism.

16.6.1 Introduction to Ethernet Link Aggregation

As broadband services are widely used, carriers require higher bandwidth and reliability for Ethernet links.

Hardware upgrades can increase Ethernet link bandwidth but at high costs. In addition, hardware upgrades are less flexible than software upgrades. Ethernet link aggregation addresses these issues and supports the following functions:

- The maximum bandwidth of a link aggregation group (LAG) is equal to the total bandwidths of all links in a LAG. Hardware upgrades are not required and costs are curtailed.
- Traffic in a LAG is distributed to member links using the load sharing algorithm, implementing load sharing and improving link usage.
- Member links in a LAG dynamically back up each other. When one link is interrupted, a backup link immediately takes over.
- Link aggregation functions between interconnected devices only and is independent of the network topology.

A logical link aggregating several physical links is called a LAG.

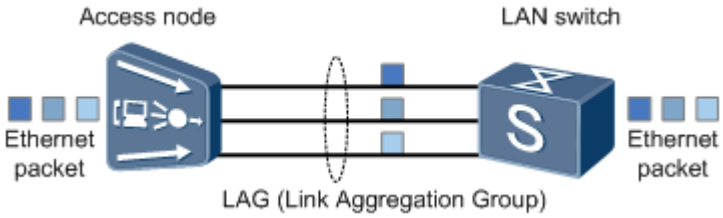
NOTE

Link aggregation is also called port aggregation because each link corresponds to two specific ports on an Ethernet network.

As shown in **Figure 16-26**, two devices are interconnected through several pairs of Ethernet ports. These ports are bound into a logical link, called a LAG.

 Copyright © Huawei Technologies Co., Ltd.

Figure 16-26 Link aggregation



16.6.2 Principle of LACP

Definition

LACP was developed based on IEEE 802.1AX. LACP dynamically aggregates and deaggregates Ethernet links using the following functions:

- Enables the data switching device to use a standard negotiation mode for link aggregation.

  After LACP is enabled, the access device aggregates multiple links according to its configuration and enables the aggregation link to transmit and receive data.

  Member ports in a LAG send LACP packets to notify the peer device of the information, including the system priority, system MAC address, port priority, port ID, and administrative key.

  After receiving the information, the peer device compares the information with the one saved on other ports to select ports that can be aggregated. The interconnected devices reach an agreement on the ports that can transmit and receive data and determine which links carry traffic.

  The LACP packets are sent in any of the following modes:

  – Event-triggered transmission

    A change in the state of the local device or in the local configuration triggers the generation and transmission of a new LACP packet.

  – Periodic transmission

    When an aggregation link is stable, the system state is periodically exchanged to maintain the aggregation link.

- Maintains the aggregation link by periodically exchanging the system state when the link is stable.

  After aggregation links are generated, LACP maintains link status. When the aggregation condition changes, LACP automatically adjusts the links in the aggregation group. The aggregation condition changes if one or more of the following occurs:

  – Physical port status changes.

  – Board status changes.

  – Results negotiated with the peer end are changed.

 Copyright © Huawei Technologies Co., Ltd.

LACP Priority

LACP priority includes system priority and port priority. Priority setting allows negotiation of aggregation information between LAGs at two ends and real-time maintenance on link status.

Priority Type	Function	Description
System priority	Specifies the priority of a LAG on the device.	Determines which party in the interconnected devices is dominant in protocol negotiation. A device with a higher priority is dominant in LACP protocol negotiation, and the information provided by it prevails for such operations as selection of an active port. **NOTE** If interconnected devices have the same system priority, compare the system MAC addresses. The device with a smaller MAC address is dominant in LACP protocol negotiation.
Port priority	Specifies the priority of a port in a LAG.	Specifies the active ports in a LAG that will carry services with preference. Active ports in a LAG can be adjusted based on port priorities. **NOTE** In a LAG, if two ports have the same priority, the port with smaller subrack ID, slot ID, and port ID (*frameid/slotid/portid*) is preferentially selected as the active one.

NOTE

System priority and port priority work together to determine which ports in a LAG are used to carry services with preference. System priority prevails over port priority.

Timeout Time for Exchanging LACP Packets

The access device periodically sends and receives LACP packets to prevent information loss. If the device does not receive any LACP packets within three LACP packet exchanging periods, the device determines that the port is faulty.

To ensure detection sensitivity, LACP defines long timeout and short timeout.

The device uses short time for exchanging LACP packets by default except that the peer device requires long timeout.

16.6.3 Introduction to Load Bearing

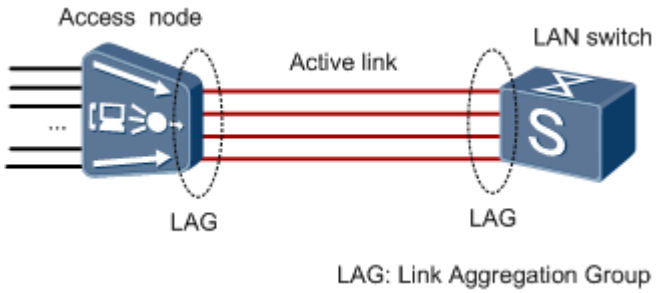
The device supports load sharing and load non-sharing for a LAG. Both load sharing and load non-sharing increase link bandwidth and improve link reliability, but they achieve this in different ways.

 Copyright © Huawei Technologies Co., Ltd.

Load Sharing Link Aggregation

Each member link in a load sharing LAG carries traffic. Member links in the LAG share the load, as shown in **Figure 16-27**.

Figure 16-27 Load Sharing link aggregation



To ensure packets carried on member links are correctly received on the peer device and ensure load balancing over member links in a LAG, the device allocates packets using hash algorithms based on:

- Source MAC addresses
- Source MAC addresses + destination MAC addresses
- Source IP addresses + destination IP addresses
- MPLS labels
- Packet quintuple (protocol, source IP address, destination IP address, source port, and destination port)

NOTE

Applications vary depending on device capabilities. Traffic distribution is calculated based on the hash algorithm. Therefore, bandwidth cannot be evenly allocated among links and different ports may differ significantly in traffic.

If member links in a LAG change or some member links become faulty, the device automatically reallocates traffic.

Load Non-Sharing Link Aggregation (Active/Standby Mode)

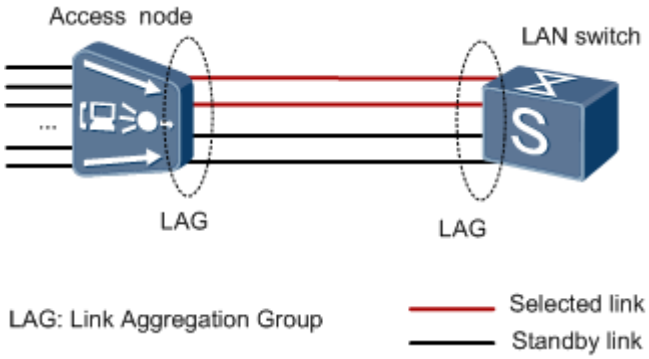
In a load non-sharing LAG, only some member links carry traffic and other links are in the standby state. Member links work in a hot standby mechanism, as shown in **Figure 16-28**.

NOTE

Load non-sharing link aggregation is implemented by configuring the maximum number of links in a LAG, and it applies to a scenario where the device is single-homed to the upstream device.

 Copyright © Huawei Technologies Co., Ltd.

Figure 16-28 Load non-sharing link aggregation (active/standby mode)



When the selected link fails, the device selects a link from the standby links to take over.

16.6.4 Introduction to LAG Type

The device supports manual aggregation and LACP aggregation for a LAG. **Table 16-7** defines them and describes their characteristics. If both interconnected devices support LACP, LACP aggregation is recommended.

Table 16-7 LAG type and characteristics

LAG Type	Definition	Application Scenario	Impact on Services
Manual aggregation	LACP is not enabled on the device. The device determines whether to aggregate a port according to its physical status (up or down), working mode, and rate.	The device is interconnected with the device that does not support LACP.	Interconnected devices do not fully negotiate with each other to aggregate links. In this case, when links work in forced GE mode and the receive or transmit direction of a member link fails (for example, when the optical fiber connected to a receive or transmit Ethernet port is cut), or when the link is incorrectly connected, the service transmit end cannot detect the fault. As a result, data may be lost.

 Copyright © Huawei Technologies Co., Ltd.

LAG Type	Definition	Application Scenario	Impact on Services
LACP aggregation	LACP is enabled on the device. By running LACP, interconnected devices have the same aggregation information, including physical port status, working mode, rate, and LACP priority. An LACP LAG has more accurate and effective control over link aggregation than a manual LAG.	The device is interconnected with the device that supports LACP.	None

Table 16-8 lists the relationship between LAG type and load sharing mode.

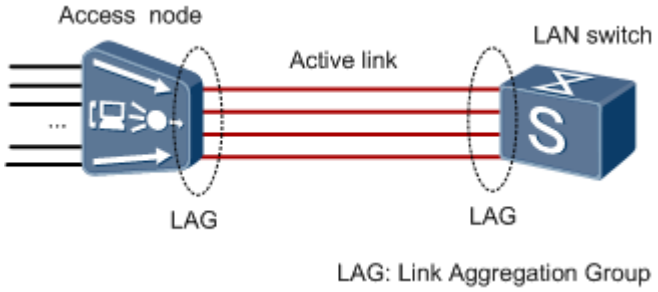
Table 16-8 Relationship between LAG type and load sharing mode

LAG Type	Load Sharing	Load Non-Sharing (Active/Standby Mode)
Manual aggregation	Supported	Not supported
LACP aggregation	Supported	Supported

Manual Aggregation

Figure 16-29 shows manual aggregation. In this type, load is shared on all member links, and there is no standby link.

Figure 16-29 Load Sharing link aggregation



 Copyright © Huawei Technologies Co., Ltd.

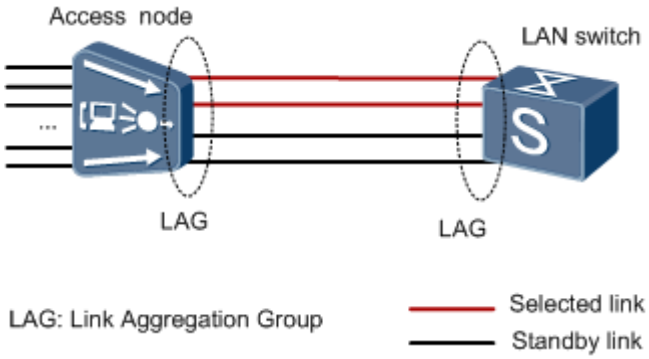
LACP Aggregation

LACP aggregation has the advantages of manual aggregation and LACP. In LACP aggregation, LACP is enabled only on the specified port, facilitating network topology control.

The network topology of LACP aggregation in load sharing is the same as that of manual aggregation, as shown in **Figure 16-29**.

Figure 16-30 shows LACP aggregation in load non-sharing.

Figure 16-30 Load non-sharing link aggregation (active/standby mode)



Member links in an LACP LAG are either selected links or standby links.

- A selected port is a working port that carries traffic.
- A standby port does not carry traffic.

The selected or standby state of a port changes with the change in system running or external environment.

- The port status change results in port status changes at the LACP layer. For example, if a port is faulty in a LAG, its state is switched to standby at the LACP layer.
- The LACPDU exchange may result in port status changes at the LACP layer. For example, a device may change the status of its ports after it receives an LACPDU from its peer device.

NOTE

For details about the frame structure of LACPDUs, see IEEE 802.1AX-2008.

16.6.5 Introduction to Port Type

Member ports in a LAG are classified into primary ports and secondary ports.

 Copyright © Huawei Technologies Co., Ltd.

Table 16-9 Port type and characteristics

Port Type	Definition	Port Characteristics	
Primary port	Logical port representing a LAG in service configuration, used in service configuration and query.	Similarity: the primary and secondary ports are defined logically.	Difference • The primary port represents the LAG to participate in service configuration whereas secondary ports cannot participate in service configuration. • A LAG has only one primary port but can have several secondary ports. • The primary port can quit its affiliated LAG only after the LAG is deleted. A secondary port can be added to or deleted from a LAG. • After a LAG is deleted, its services are still carried by the primary port.
Secondary port	All member ports except for the primary port.		

NOTE

- An Ethernet port can be added to only one LAG.
- If the primary port is faulty, secondary ports work. However, service-related operations can only be performed on the primary port.
- The device supports a LAG containing only one port. When the LAG requires expansion, you only need to add ports to the LAG but do not need to modify services.
- When a service is configured on ports in an inter-board LAG group, the service can be successfully configured only if all ports in this group support this service. If a port in the group does not support this service, the system displays message "The port does not support this operation" or "The board does not support this operation".

16.6.6 Principle of LAG Setup Process

NOTE

A manual LAG does not use LACP. Therefore, principles of a manual LAG are not described here. For details, see IEEE 802.1AX.

The process of setting up an LACP LAG is as follows:

1. Interconnected devices send LACPDUs to each other.
2. Interconnected devices determine the actor (indicating the party that is dominant in negotiation) based on LACP system priorities and system IDs.

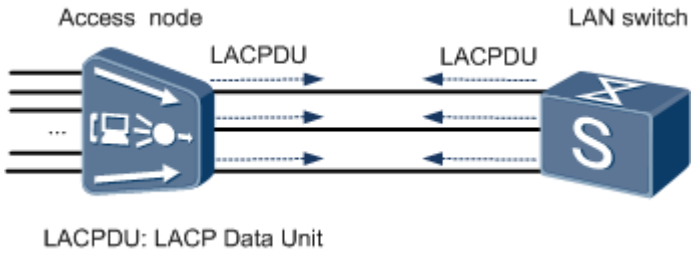
 Copyright © Huawei Technologies Co., Ltd.

3. Interconnected devices determine active ports (ports carrying traffic) based on the port LACP priorities and port IDs of the Actor.

Exchanging LACPDUs Between Interconnected Devices

An LACP LAG is set up on interconnected devices, and member ports are added to the LAG. The member ports are enabled with LACP, enabling interconnected devices to exchange LACPDUs to each other, as shown in **Figure 16-31**.

Figure 16-31 Exchanging LACPDUs between interconnected devices of an LACP LAG

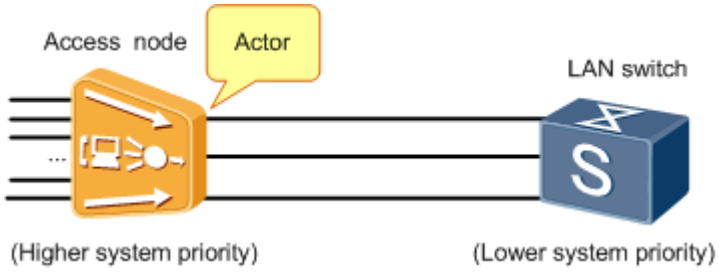


Determining the Actor

Interconnected devices receive LACPDUs from each other. Use the LAN switch shown in **Figure 16-32** as an example to describe the process. After receiving LACPDUs from the access node, the LAN switch checks and records information about the access node and compares system priorities. If the system priority of the access node is higher than that of the LAN switch, the access node acts as the actor.

If the system priority of the access node is the same as that of the LAN switch, the party with a smaller MAC address in the system ID functions as the actor.

Figure 16-32 Determining the actor



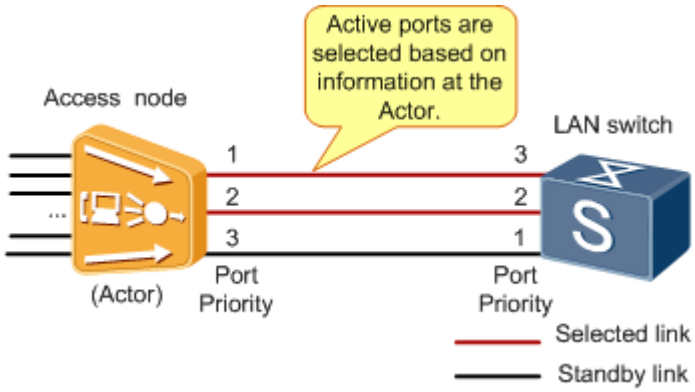
Selecting Active Ports

Interconnected devices select active ports based on the priorities of ports on the actor after the actor is selected.

After the same active port is selected, selected links in the LAG are determined, and traffic is distributed in load sharing across the links, as shown in **Figure 16-33**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 16-33 Selecting active ports



16.6.7 Principle of LAG Link Switchover and Reversion

Switching Condition

In an LACP LAG, a link switchover is triggered if a device at one end detects one of the following events:

- A selected link goes Down.
- LACP detects a link failure.
- An active port becomes unavailable.

Switchover Process

When any of the preceding trigger conditions is met, the link switchover is performed in one of the following processes:

For load non-sharing LAGs:

1. The faulty link is disabled.
2. The standby link with the highest priority is selected to replace the faulty selected link.
3. The standby link with the highest priority becomes the selected link and then forwards data.

For load sharing LAGs:

1. The faulty link is disabled.
2. Traffic is reallocated to member links using the load sharing algorithm.

Switchover Reversion Process

If an LACP LAG works in load sharing mode and the active port before switchover recovers, the original faulty link is enabled, and traffic is reallocated using the load sharing algorithm.

If an LACP LAG works in load non-sharing mode, you can set the link revertive mode to revertive or non-revertive.

 Copyright © Huawei Technologies Co., Ltd.

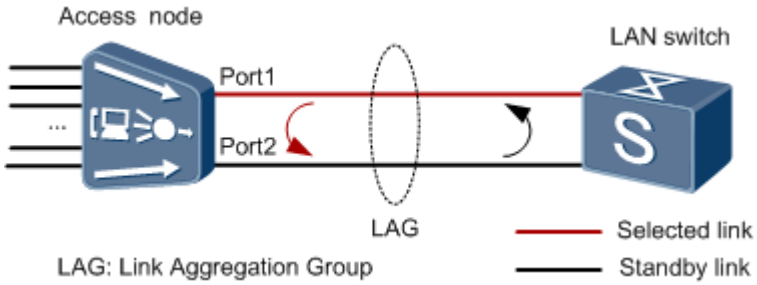
- When an LACP load non-sharing LAG is set to revertive, services are switched back to the link of a higher priority after this link recovers, as shown in **Figure 16-34**.
- When an LACP load non-sharing LAG is set to non-revertive, services are not switched back to the selected link after this link recovers, but are still transmitted on the current link.

As shown in **Figure 16-34**, ports 1 and 2 are member ports in an LACP LAG, port 1 is the active port (working in the selected link), and port 2 is the inactive port (working in the standby link).

After port 1 fails, port 2 takes over as the active port.

After port 1 recovers, port 1 works as an active port, and port 2 still works as an inactive port.

Figure 16-34 Reversing a switchover (LACP LAG)



16.6.8 Configuring Ethernet Link Aggregation

Configure Ethernet link aggregation to increase link bandwidth and improve link reliability, without performing a hardware upgrade.

Prerequisites

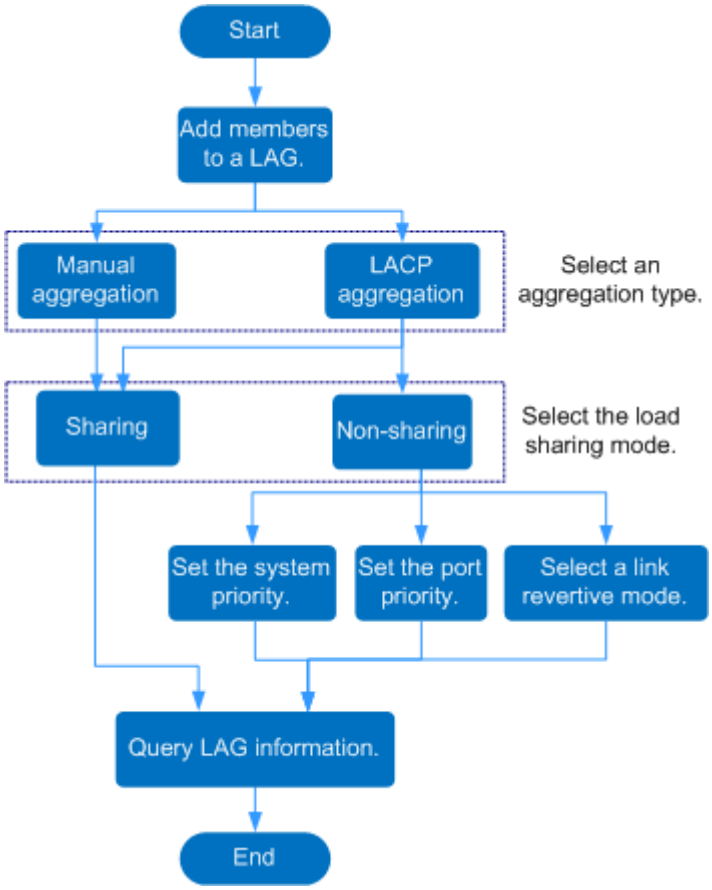
- Interconnected devices, hardware, and port attributes must support LAGs.
- The member ports of LAG do not have static MAC addresses. You can run the **display mac-address** command to query whether an aggregated port has static MAC address.

Context

Figure 16-35 shows the flowchart for configuring a LAG.

 Copyright © Huawei Technologies Co., Ltd.

Figure 16-35 Configuration flowchart



Procedure

Step 1 Create a LAG and select the aggregation type.

Run the **link-aggregation** command to add multiple upstream Ethernet ports to the same LAG to implement protection and load sharing between ports. The port with the smallest port ID is the master port.

NOTE

If the device is interconnected with the device that supports LACP, static aggregation is recommended. If the device is interconnected with the device that does not support LACP, only manual aggregation can be used.

Step 2 (Optional) Add a LAG member port.

Perform this step when the LAG bandwidth or link reliability needs to be improved further. Run the **link-aggregation add-member** command to add an Ethernet port to an existing LAG to increase the LAG bandwidth and improve the link reliability.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

If the port to be added to or deleted from a LAG is connected to the peer device, run the **shutdown(Ethernet)** command to deactivate the Ethernet port or remove the optical fiber to prevent loops.

Step 3 (Optional) Select the load carrying type.

NOTE

If the load sharing type is not configured, a LAG works in load sharing mode by default.

This step is required only when a static LAG is configured.

Configuring the maximum selected links in a LAG implements traffic allocation in load non-sharing mode. For example, M+N links are configured in a LAG. Then, run the link-aggregation max-link-number command to specify N selected links. The remaining M links are standby ones. If a selected link is disconnected, a standby link automatically changes to the selected one.

Step 4 (Optional) Set the system priority and port priority.

This step is required only when a static LAG is configured.

- LACP system priority: If the access device is dual homed to two convergence devices, the access device determines the selected and standby LAGs. Run the **lacp priority system** command to set the LACP system priority of the access device to be higher than that of the peer device.
- LACP port priority: LACP port priority must be used together with the maximum number of links. If a port is required preferentially for carrying services, set its priority higher. Run the **lacp priority port** command to change the link priority so that the standby link and the selected link can be switched over.

Step 5 (Optional) Selected the link revertive mode.

When a static LAG in load non-sharing mode is configured, run the **lacp preempt** command to set whether traffic is switched back to the original link if the link failure is rectified.

Step 6 (Optional) Query LAG information.

Run the **display link-aggregation** command to query the LAG information, including primary port, number of links, aggregation type (manual or static), and maximum number of links.

----End

Example

Assume the following configurations: The MA5800 transmits services upstream using the upstream ports, upstream ports0/9/0 and 0/9/1 are configured in an upstream port LAG, packets are distributed to the LAG member ports according to the source MAC address, and the working mode is LACP static aggregation. To perform these configurations, run the following commands:

```
huawei(config)#link-aggregation 0/9 0-1 ingress workmode lacp-static
huawei(config)#display link-aggregation all
 ------------------------------------------------------------------------------
```

 Copyright © Huawei Technologies Co., Ltd.

```
Master port  Link aggregation mode  Port NUM  Work mode  Max link number
------------------------------------------------------------------------------
0/9/0        ingress                2  lacp-static             -
------------------------------------------------------------------------------
Total: 1 link aggregation(s)
```

16.6.9 Ethernet Link Aggregation Reference Standards and Protocols Compliance

Ethernet link aggregation complies with IEEE 802.1AX-2008, IEEE standard for local and metropolitan area networks (link aggregation).

16.7 MSTP

The Multiple Spanning Tree Protocol (MSTP) is compatible with STP and RSTP.

16.7.1 Introduction to MSTP

Definition

The Spanning Tree Protocol (STP) applies to a loop network to realize path redundancy through certain algorithms. STP also prunes a loop network into a loop-free tree network. This helps to avoid proliferation and infinite loop of packets in the loop network.

The Rapid Spanning Tree Protocol (RSTP) is an improvement on STP. The rapidness of RSTP relies on the greatly shortened delay for the designated port and the root port to turn into the forwarding state in a certain condition. For details, see "Principle of RSTP" in "**16.7.2 Principle of MSTP**." This helps to shorten the time for stabilizing the network topology.

The Multiple Spanning Tree Protocol (MSTP) is compatible with STP and RSTP.

Purpose

Although STP can prune a loop network into a loop-free network, it fails to transit fast. Even a port in a point-to-point link or an edge port has to wait double Forward Delay time before it can turn into the forwarding state.

RSTP features fast convergence; however, like STP, RSTP still has the following defects:

- All the bridges in a local area network (LAN) share a same spanning tree, and fail to block redundant links by VLAN.
- The packets of all the VLANs are forwarded along the same spanning tree. Therefore, load sharing of data traffic cannot be implemented between VLANs.

MSTP can be a remedy to the defects of STP and RSTP. It not only realizes fast convergence, but also enables traffic of different VLANs to be forwarded along their respective paths. This helps to provide a better load sharing mechanism for redundant links.

MSTP sets VLAN mapping tables (relationship tables between VLANs and spanning trees) to associate VLANs and spanning trees. MSTP divides a switching

 Copyright © Huawei Technologies Co., Ltd.

network into multiple regions. Each region contains multiple spanning trees, and each spanning tree is independent from others.

MSTP prunes a loop network to a loop-free tree network to avoid proliferation and infinite loop of packets in the loop network. It also provides multiple redundant paths for data forwarding to realize load sharing of VLAN data during forwarding.

16.7.2 Principle of MSTP

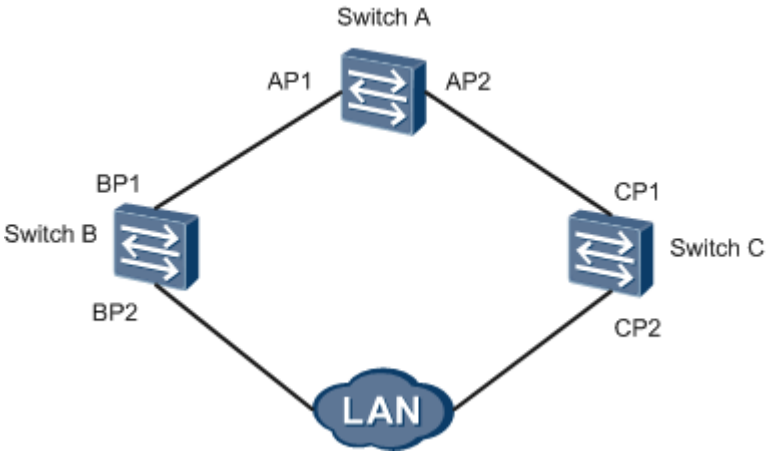
Principle of STP

STP determines the topology of a network by transmitting a certain special message (configuration message as defined in IEEE 802.1D) between bridges. A configuration message contains sufficient information to enable the bridge to complete the calculation of the spanning tree.

The following defines the designated port and the designated bridge:

- For a bridge (such as bridge A), the designated bridge is a bridge that is directly connected to bridge A and forwards data packets to bridge A. The designated port is the port in the designated bridge through which the data packets are forwarded to bridge A.
- For a LAN, the designated bridge is a bridge that forwards data packets to the LAN. The designated port is the port in the designated bridge through which the data packets are forwarded to the LAN.

Figure 16-36 Schematic drawing of designated bridge and designated port



As shown in **Figure 16-36**:

- AP1, AP2, BP1, BP2, CP1, and CP2 are ports in Switch A, Switch B, and Switch C respectively.
- Switch A forwards data to Switch B through port AP1, and then the designated bridge of Switch B is Switch A, and the designated port is port AP1 in Switch A.

 Copyright © Huawei Technologies Co., Ltd.

- Switch B and Switch C are connected to the LAN. If Switch B forwards data packets to the LAN, the designated bridge of the LAN is Switch B, and the designated port is port BP2 in Switch B.

In STP, the configuration message is forwarded as follows:

1. In network initialization, all the bridges work as the root bridge of the spanning tree.
2. The designated port of a bridge takes the hello time as the interval for sending its configuration messages. If the port that receives the configuration message is a root port, the bridge increases the message age contained in the configuration message by degrees and enables the timer to time the configuration message.
3. If a path fails, the root port on this path receives new configuration messages no longer, and the old configuration messages are discarded due to timeout. This results in recalculation of the spanning tree. A new path then is created to replace the faulty path and recover the network connectivity.

The new configuration message upon the recalculation, however, will not immediately spread throughout the entire network. In this case, the old root port and designated port that fail to discover the topology change will forward their data along the old paths. If the selected root port and designated port forward data immediately, a temporary loop may be created.

Therefore, STP adopts a state transition mechanism. That is, the root port and the designated port have to experience a transition state before they can re-forward data. The transition state turns into the forwarding state upon Forward Delay. This delay guarantees that the new configuration message has spread throughout the entire network.

Defects of STP

- In case of topology change or link failure, a port has to wait double Forward Delay time before it can turn from the blocking state to the forwarding state. Therefore, in case of topology change, double Forward Delay time (at least scores of seconds) is required to restore the network connectivity.
- The entire bridged LAN uses a single spanning tree instance. Therefore, when the network is large, a longer convergence time may be required or the topology changes frequently.

Principle of RSTP

RSTP is an improvement on STP. The rapidness of RSTP relies on the greatly shortened delay for the designated port and the root port to turn into the forwarding state in a certain condition. This helps to shorten the time for stabilizing the network topology.

In comparison with STP, RSTP improves in the following aspects:

- First improvement:
  - The alternate port and backup port are set for rapid switching of the root port and designated port.
  - When the root port fails, the alternate port quickly switches to the new root port and turns into the forwarding state without delay.

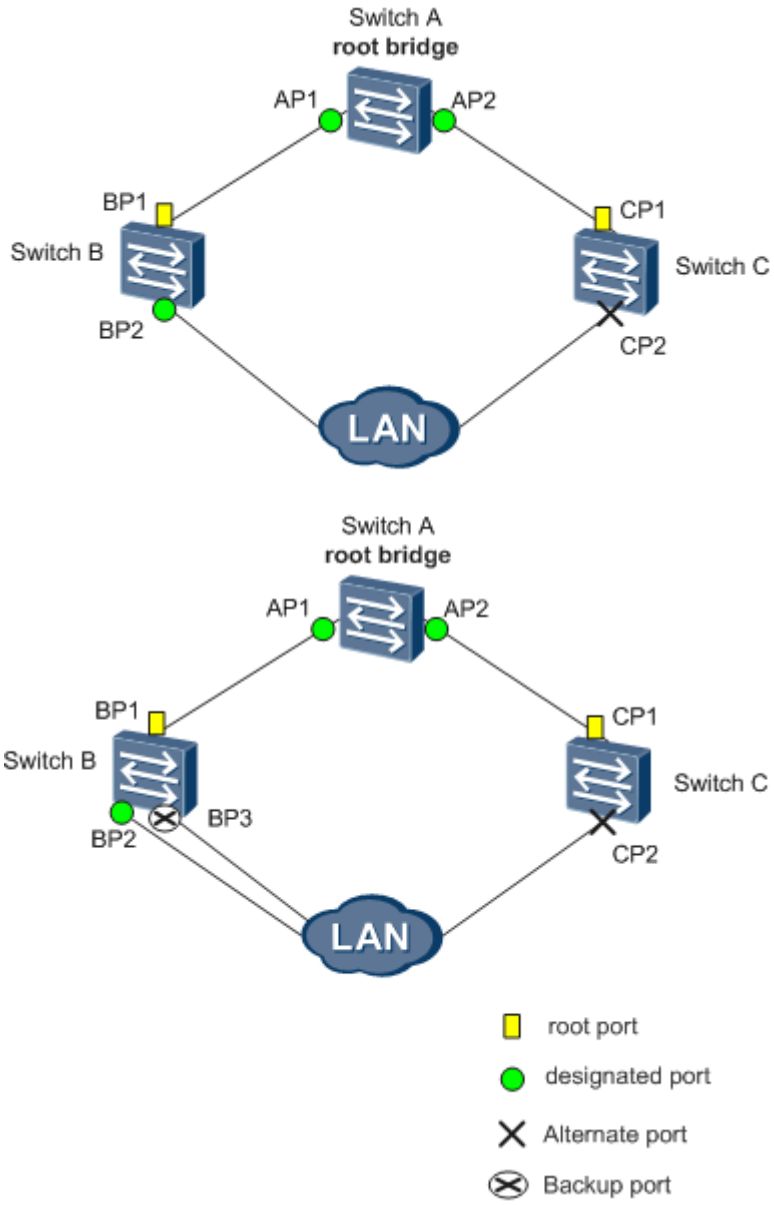
 Copyright © Huawei Technologies Co., Ltd.

  - When the designated port fails, the backup port quickly switches to the new designated port and turns into the forwarding state without delay.
- Second improvement:
  - In a point-to-point link connected with two switching ports, a designated port turns into the forwarding state without delay after one handshake with the downstream bridge.
  - In a shared link connected with at least three bridges, the downstream bridge does not respond to the handshake request sent from the upstream designated port, and the designated port has to wait double Forward Delay time before it turns into the forwarding state.
- Third improvement:
  - A port that is directly connected to a terminal and is not connected to any other bridge is defined as an edge port. The edge port can directly turn into the forwarding state without delay.
  - Because a bridge does not know whether a port is directly connected to a terminal, the edge port must be configured manually.

RSTP defines four port roles: root port, designated port, alternate port, and backup port. As shown in **Figure 16-37**.

 Copyright © Huawei Technologies Co., Ltd.

Figure 16-37 Diagram of port roles



The functions of the root port and designated port are the same as those defined in STP. The alternate port and backup port are described as follows:

- From the perspective of configuration BPDU transmission:
  - An alternate port is blocked after learning the configuration BPDUs sent by other bridges.
  - A backup port is blocked after learning the configuration BPDUs sent by itself.
- From the perspective of user traffic
  - An alternate port backs up the root port and provides an alternate path from the designated bridge to the root bridge.
  - A backup port backs up the designated port and provides an alternate path from the root node to the leaf node.

 Copyright © Huawei Technologies Co., Ltd.

After all RSTP-capable ports are assigned roles, topology convergence is completed.

The bridges that adopt RSTP are compatible with the bridges which adopt STP. The bridges that adopt RSTP can identify both STP and RSTP packets and apply them to calculation of the spanning tree.

Defects of RSTP

Although RSTP features fast convergence, like STP, RSTP still has the following defects: All the bridges in a LAN share a same spanning tree, and thus the packets of all the VLANs cannot be forwarded equally. Furthermore, the packets of some VLANs cannot be forwarded.

Principle of MSTP

MSTP can compensate for the defects of STP and RSTP. It not only realizes fast convergence, but also enables traffic of different VLANs to be forwarded along their respective paths. This helps to provide a better load sharing mechanism for redundant links.

MSTP sets VLAN mapping tables (relationship tables between VLANs and spanning trees) to associate VLANs and spanning trees. MSTP divides a switching network into multiple regions. Each region contains multiple spanning trees, and each spanning tree is independent from others. Multiple spanning trees can run on each bridge to forward the packets of different VLANs.

MSTP divides the entire Layer 2 network into multiple spanning tree (MST) regions. These regions and the other bridges and LANs are connected into a single common spanning tree (CST). Multiple spanning trees are created in a region through calculation. Each spanning tree is defined as a multiple spanning tree instance (MSTI). MSTI 0 is defined as an internal spanning tree (IST). MSTP connects all bridges and LANs with a single common and internal spanning tree (CIST) which consists of the CST and the IST. Like RSTP, MSTP calculates the spanning tree according to the configuration message. The configuration message, however, contains the message of MSTP on the bridge.

- Calculation of CIST
  - Select a bridge with the highest priority within the entire network as the CIST root by comparing the configuration messages.
  - In each MST region, MSTP creates an IST through calculation. Meanwhile, MSTP regards each MST region as a single bridge, and then creates a CST between regions.
  - The CST and IST form the CIST that connects all the bridges in a bridge network.
- Calculation of MSTI

  In an MST region, MSTP creates different MSTIs for different VLANs according to the mapping between the VLANs and the spanning tree instances. Each spanning tree is calculated independently. The process is similar to that in which the RSTP calculates the spanning tree.

Based on RSTP, MSTP has two additional port types. MSTP ports can be root ports, designated ports, alternate ports, backup ports, edge ports, master ports, and regional edge port.

 Copyright © Huawei Technologies Co., Ltd.

The functions of root ports, designated ports, alternate ports, backup ports, and edge ports have been defined in RSTP. **Table 16-10** lists all port roles in MSTP.

NOTE

Except edge ports, all ports participate in MSTP calculation.

A port can play different roles in different spanning tree instances.

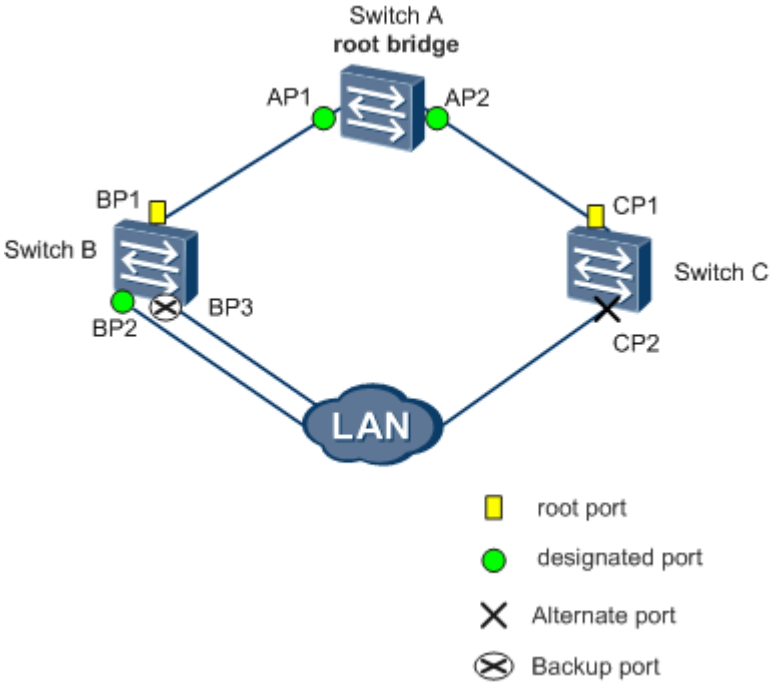
Table 16-10 Port roles

Port Role	Description
Root port	A root port is the non-root bridge port closest to the root bridge. Root bridges do not have root ports. Root ports are responsible for sending data to root bridges. As shown in **Figure 16-38**, Switch A is the root; BP1 is the root port on Switch B; CP1 is the root port on Switch C.
Designated port	The designated port on a switching device forwards BPDUs to the downstream switching device. As shown in **Figure 16-38**, AP1 and AP2 are designated ports on Switch A; BP2 is a designated port on Switch B.
Alternate port	• From the perspective of sending BPDUs, an alternate port is blocked after a BPDU sent by another bridge is received. • From the perspective of user traffic, an alternate port provides an alternate path to the root bridge. This path is different than using the root port. As shown in **Figure 16-38**, CP2 is an alternate port.
Backup port	• From the perspective of sending BPDUs, a backup port is blocked after a BPDU sent by itself is received. • From the perspective of user traffic, a backup port provides a backup/redundant path to a segment where a designated port already connects. As shown in **Figure 16-38**, BP3 is a backup port.
Master port	A master port is on the shortest path connecting MST regions to the CIST root. BPDUs of an MST region are sent to the CIST root through the master port. Master ports are special regional edge ports, functioning as root ports on ISTs or CISTs and master ports in instances. As shown in **Figure 16-39**, Switch A, Switch B, Switch C, and Switch D form an MST region. AP1 on Switch A, being the nearest port in the region to the CIST root, is the master port.

 Copyright © Huawei Technologies Co., Ltd.

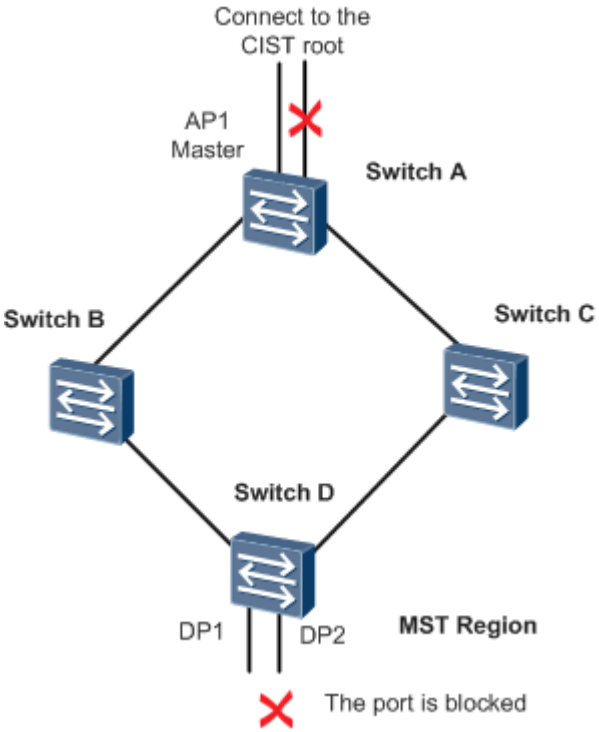
Port Role	Description
Regional edge port	A regional edge port is located at the edge of an MST region and connects to another MST region or an SST. During MSTP calculation, the roles of a regional edge port in the MSTI and the CIST instance are the same. If the regional edge port is the master port in the CIST instance, it is the master port in all the MSTIs in the region. As shown in **Figure 16-39**, AP1, DP1, and DP2 in an MST region are directly connected to other regions, and therefore they are all regional edge ports of the MST region. AP1 is a master port in the CIST. Therefore, AP1 is the master port in every MSTI in the MST region.
Edge port	An edge port is located at the edge of an MST region and does not connect to any switching device. Generally, edge ports are directly connected to terminals.

Figure 16-38 Root port, designated port, alternate port, and backup port



 Copyright © Huawei Technologies Co., Ltd.

Figure 16-39 Master port and regional edge port



Implementation of MSTP on the MA5800

MSTP is compatible with STP and RSTP. The bridges that adopt MSTP can identify both STP and RSTP packets and apply them to calculate spanning tree.

Besides the basic functions of MSTP, the MA5800 provides some special functions, such as:

- BPDU protection

  For an access device, the access port is generally connected to a terminal (such as a PC) or file server. In this case, the access port is set to an edge port for the purpose of fast transition. When receiving a configuration message (BPDU), the edge port switches to a non-edge port automatically, the spanning tree is re-calculated and the topology changes accordingly.

  In normal conditions, an edge port cannot receive STP configuration messages. If the bridge is maliciously attacked by forged configuration messages, the network will be attacked. The BPDU protection function can prevent such network attacks.

  After the BPDU protection function is enabled on the MA5800, if an edge port receives a configuration message, the system shuts down the edge port, and notifies the network management system of the related information. Only network administrators can enable the port that is shut down.

  It is recommended that you enable the BPDU protection function on the MA5800 which is configured with an edge port.

- Root protection

 Copyright © Huawei Technologies Co., Ltd.

Because of wrong configurations by the maintenance personnel or malicious network attacks, a legal root bridge in the network may receive a configuration message with a higher priority. In this case, this root bridge may become a non-root bridge and the topology changes accordingly. Such illegal change results in transfer of traffic in high-speed links to low-speed links, thus causing network congestion.

The root protection function is a solution to this problem.

When the root protection function is enabled for a port, the port is always a designated port. If the port receives a configuration message with a higher priority, and is to become a non-designated port, the port will turn into the listening state and will not forward packets (that is, the link connected to the port is disconnected). If the port does not receive a configuration message of a much higher priority within a certain long period of time, the port will turn into the normal state.

- Loop protection

  A bridge maintains the states of the root port and other blocked ports by continuously receiving BPDUs from the upstream bridge.

  In case of link congestion or failure, these ports fail to receive BPDUs from the upstream bridge. For this reason, the bridge will re-select its root bridge. The previous root bridge switches to the designated port, and the blocked ports turn to the forwarding state. As a result, loops are created in the switching network.

  The loop protection function is a solution to this problem.

  After receiving the BPDUs (excluding the TCN packets) again, a port under loop protection normally processes the packets, selects the role, and resets the forwarding state of the port. The port is not always in the blocked state.

  When the loop protection function is enabled, if the root port switches to a non-root port, it will turn into the discarding state, and the blocked ports will remain in the discarding state. Therefore, no packets are forwarded, and no loop is created in the network.

NOTE

The three protection functions conflict with each other.

16.7.3 Configuring the MSTP

The MA5800 supports the application of the Multiple Spanning Tree Protocol (MSTP), Spanning Tree Protocol (STP), and Rapid Spanning Tree Protocol (RSTP). The MA5800 supports the MSTP ring network, which can meet various networking requirements.

Context

- MSTP applies to a redundant network. It makes up for the drawback of STP and RSTP. MSTP makes the network converge fast and the traffic of different VLANs distributed along their respective paths, which provides a better load-sharing mechanism.
- MSTP trims a loop network into a loop-free tree network. It prevents the proliferation and infinite cycling of the packets in the loop network. In addition, MSTP supports load sharing by VLAN during data transmission.

 Copyright © Huawei Technologies Co., Ltd.

- The status of the transparent transmission for bridge protocol data unit (BPDU) packets configured by running the **bpdu tunnel** command determines the destination MAC address of MSTP packets.
  - If the transparent transmission of BPDU packets is disabled in all VLANs, the destination MAC address of MSTP packet is 00:E0:FC:00:00:00.
  - If the transparent transmission of BPDU packets is enabled in any VLAN, the destination MAC address of MSTP packet is 00:E0:FC:00:00:08.

Procedure

Step 1 Enabling the MSTP function.

- By default, the MSTP function is disabled.
- After the MSTP function is enabled, the device determines whether it works in STP compatible mode or MSTP mode based on the configured protocol.
- After the MSTP function is enabled, MSTP maintains dynamically the spanning tree of the VLAN based on the received BPDU packets. After the MSTP function is disabled, the MA5800 becomes a transparent bridge and does not maintain the spanning tree.

1. Run the **stp enable** command to enable the MSTP function of the bridge.
2. Run the **stp port enable** command to enable the MSTP function of the port.
3. Run the **display stp** command or the **display stp port** command to query the MSTP state of the bridge or the port.

Step 2 Configuring the MST region name.

1. Run the **stp region-configuration** command to enter MST region mode.
2. Run the **region-name** command to configure the name of the MST region.

   By default, the MST region name is the bridge MAC address of the device.

Step 3 Configuring the MSTP instance.

The MSTP protocol configures the VLAN mapping table (mapping between the VLAN and the spanning tree), which maps the VLAN to the spanning tree.

1. Run the **stp region-configuration** command to switch over to MST region mode.
2. Run the **instance vlan** command to map the specified VLAN to the specified MSTP instance.
   - By default, all VLANs are mapped to CIST, that is, instance 0.
   - One VLAN can be mapped to only one instance. If you re-map a VLAN to another instance, the original mapping is disabled.
   - A maximum of 10 VLAN sections can be configured for an MSTP instance.

   NOTE

   A VLAN section refers to the consecutive VLAN IDs from the start VLAN ID to the end VLAN ID.
3. Run the **check region-configuration** command to query the parameters of the current MST region.

Step 4 Activating the configuration of the MST region.

 Copyright © Huawei Technologies Co., Ltd.

1. Run the **stp region-configuration** command to switch over to MST region mode.
2. Run the **active region-configuration** command to activate the configuration of the MST region.
3. Run the **display stp region-configuration** command to query the effective configuration of the MST region.

Step 5 Setting the priority of the device in the specified spanning tree instance.

1. Run the **stp priority** command to set the priority of the device in the specified spanning tree instance.
2. Run the **display stp** command to query the MSTP configuration of the device.

Step 6 Other optional configurations.

- Setting the MST region parameters.
  - Run the **stp md5-key** command to set the MD5-Key for the MD5 encryption algorithm configured on the MST region. MD5 is an insecure encryption algorithm. Therefore, using an algorithm with a higher security level is recommended.
  - In the MSTP region mode, run the **vlan-mapping module** command to map all VLANs to the MSTP instances by modular arithmetic.
  - In the MSTP region mode, run the **revision-level** command to set the MSTP revision level of the device.
  - Run the **reset stp region-configuration** command to restore the default settings to all parameters of the MST region.
- Specifying the device as a root bridge or a backup root bridge.
  - Run the **stp root** command to specify the device as a root bridge or a backup root bridge.
- Setting the time parameters of the specified network bridge.
  - Run the **stp timer forward-delay** command to set the Forward Delay of the specified network bridge.
  - Run the **stp timer hello** command to set the Hello Time of the specified network bridge.
  - Run the **stp timer max-age** command to set the Max Age of the specified network bridge.
  - Run the **stp time-factor** command to set the timeout time factor of the specified network bridge.
- Setting the parameters of the specified port.
  - Run the **stp port transmit-limit** command to set the number of packets transmitted by the port within the Hello Time.
  - Run the **stp port edged-port enable** command to set the port as an edge port.
  - Run the **stp port cost** command to set the path cost of a specified port.
  - Run the **stp port port-priority** command to set the priority of the specified port.
  - Run the **stp port point-to-point** command to set whether the link that is connected to the port is a point-to-point link.

 Copyright © Huawei Technologies Co., Ltd.

- Configuring the device protection function.
  - Run the **stp bpdu-protection enable** command to enable the BPDU protection function of the device.
  - Run the **stp port loop-protection enable** command to enable the loop protection function of the port.
  - Run the **stp port root-protection enable** command to enable the root protection function of the port.
- Setting the maximum number of hops of the MST region.
  - Run the **stp max-hops** command to set the maximum number of hops of the MST region.
- Setting the diameter of the switching fabric.
  - Run the **stp bridge-diameter** command to set the diameter of the switching fabric.
- Setting the calculation standard for the path cost.
  - Run the **stp pathcost-standard** command to set the calculation standard for the path cost.
- Clear the MSTP protocol statistics.
  - Run the **reset stp statistics** command to clear the MSTP protocol statistics.

----End

Example

Configure the MSTP parameters as follows:

- Enable the MSTP function.
- Enable the MSTP function on port 0/9/0.
- Set the MSTP running mode to MSTP compatible mode.
- Configure MST region parameters:
  - Configure the MD5-Key for the MD5 encryption algorithm to 0x11ed224466.
  - Configure the MST region name to test-mstp-bridge.
  - Map VLAN2-VLAN10 and VLAN12-VLAN16 to MSTP instance 3.
  - Map all the VLANs to the specified MSTP instances using module 16.
  - Configure the MSTP revision level of the device to 100.
- Configure the maximum hops for the MST region to 10.
- Activate the configuration of the MST region manually.
- Configure the priority of the device in spanning tree instance 2 to 4096.
- Configure the current device as the root bridge of MSTP instance 2.
- Configure the diameter of the switching network to 6.
- Configure the calculation standard for the path cost to IEEE 802.1t.
- Configure the time parameters of a specified bridge:
  - Configure the forward delay to 2000 centiseconds.
  - Configure the hello time to 1000 centiseconds.

 Copyright © Huawei Technologies Co., Ltd.

  - Configure the max age to 3000 centiseconds.
  - Configure the timeout time factor to 6.
- Configure the parameters of a specified port:
  - Configure the maximum number of packets transmitted in a hello time period to 16.
  - Configure port 0/9/0 to be an edge port.
  - Configure the path cost of the port in a specified spanning tree instance to 1024.
  - Configure the priority of the port to 64.
  - The link connected to port 0/9/0 is a point-to-point link.
- Enable the BPDU protection function on the device.

```
huawei(config)#stp enable
  Change global stp state may active region configuration,it may take several
minutes,are you sure to change global stp state? [Y/N][N]y
huawei(config)#stp port 0/9/0 enable
huawei(config)#stp mode mstp
huawei(config)#stp md5-key 11ed224466
huawei(config)#stp region-configuration
huawei(stp-region-configuration)#region-name test-mstp-bridge
huawei(stp-region-configuration)#instance 3 vlan 2 to 10 12 to 16
huawei(stp-region-configuration)#vlan-mapping module 16
huawei(stp-region-configuration)#revision-level 100
huawei(stp-region-configuration)#active region-configuration
huawei(stp-region-configuration)#quit
huawei(config)#stp instance 2 priority 4096
huawei(config)#stp instance 2 root primary
huawei(config)#stp max-hops 10
huawei(config)#stp bridge-diameter 6
huawei(config)#stp pathcost-standard dot1t
huawei(config)#stp timer forward-delay 2000
huawei(config)#stp timer hello 1000
huawei(config)#stp timer max-age 3000
huawei(config)#stp time-factor 6
huawei(config)#stp port 0/9/0 transmit-limit 16
huawei(config)#stp port 0/9/0 edged-port enable
huawei(config)#stp port 0/9/0 instance 0 cost 1024
huawei(config)#stp port 0/9/0 instance 0 port-priority 64
huawei(config)#stp port 0/9/0 point-to-point force-true
huawei(config)#stp bpdu-protection enable
```

16.7.4 MSTP Reference Standards and Protocols

The following lists the reference documents of MSTP:

- IEEE Std 802.1d, 1998 Edition, Spanning Tree Protocol
- IEEE Std 802.1w-2001, Rapid Spanning Tree Protocol
- IEEE Std 802.1s-2002, Multiple Spanning Tree Protocol

16.8 ERPS

Ethernet ring protection switching (ERPS) is a ring network protocol defined in Recommendation ITU-T G.8032. Interoperation is achieved if all devices participating in the ring network support ERPS. On a network composed of ERPS-supporting devices, the service protection solution with quick switching can be implemented using a small number of links.

 Copyright © Huawei Technologies Co., Ltd.

16.8.1 Introduction to ERPS

To resolve loop issues, Huawei has supported equipment supporting the Spanning Tree Protocol (STP) and Rapid Ring Protection Protocol (RRPP). The following challenges need to be addressed:

- As higher requirements are posed on the switching time for Layer 2 Ethernet, STP supports link recovery on a basis of seconds, falling short of the carrier-class convergence performance requirements.
- RRPP is a proprietary protocol. Interoperation is not supported if a ring network is composed of devices of multiple vendors.

Ethernet ring protection switching (ERPS) is a ring network protocol defined in Recommendation ITU-T G.8032. Interoperation is achieved if all devices participating in the ring network support ERPS. In addition, ERPS supports fast convergence, meeting carrier-class reliability requirements.

- V1 supports only single-ring protection, single Ethernet ring protection (ERP) instance, and revertive switching.
- Based on V1, V2 supports multiple-ring protection, multiple ERP instances, and non-revertive switching. V2 also optimizes the flushing mechanism for the filtering database (FDB), and supports commands for a forced switchover, manual switchover, or a clearing operation.

The MA5800 implements functions of the V1 protocol using the V2 state machine, supports V1 and V2 message reception and V1 message transmission, and supports all functions of V1 and the multiple instances defined by V2.

Recommendation ITU-T G.8032 only protects MAC address+VLAN services; the MA5800 extends the protocol by protecting multicast and Layer 3 services.

The following explains the concepts involved in the description above.

Concept	Description
Single ring	An ERPS ring physically composed of a set of devices that are connected to form a closed loop.
Multiple rings	A complex network composed of multiple interconnected single rings. The single rings may be interconnected through an interconnection point or through dual interconnection nodes.
Single instance	A mechanism that allows only one logical ERPS ring to run on a physical ring network. All services are protected by this ring.

 Copyright © Huawei Technologies Co., Ltd.

Concept	Description
Multiple instances	A mechanism that implements protection switching for ERPS rings based on Multiple Spanning Tree Protocol (MSTP) instances. Multiple logical ERPS rings can run on one physical ring network. The topologies of the rings are independent of each other. These rings implement ring automatic protection switching (R-APS) protocol message exchange, fault processing, and service switchover. Therefore, each of the rings must protect a specific type of service. When the network runs in normal conditions, each ERPS ring can select its own blocking point to load balance service traffic over the rings. The selection can be implemented through configuration.
Revertive switching	A mechanism that blocks the ring protection link (RPL) and reverts to the original faulty link to carry service traffic when the fault causing the switch is cleared and the wait-to-restore (WTR) timer expires.
Non-revertive switching	A mechanism that still blocks the original faulty link after the fault causing the switch is cleared.
Forced switching	A port configured with forced switching is immediately blocked, regardless of whether other links on the ring have encountered a fault.
Manual switching	A port configured with manual switching is blocked only when the ring does not contain a faulty link or a port configured with forced switching.
clear command	A command that clears the forced switching and manual switching locally configured.

16.8.2 Basic Concepts of ERPS

This topic describes basic concepts related to the Ethernet ring protection switching (ERPS) feature to help you understand the working principles of the ERPS ring.

ERPS Ring

An ERPS ring is composed of a group of interconnected Layer 2 switches that are configured with the same control VLAN. An ERPS ring is the basic element for implementing the ITU-T G.8032 protocol. A Layer 2 switch participating in the ERPS ring is called a node. On the node, the ports that participate in the ERPS ring are called ring ports.

Port Roles

The ITU-T G.8032 protocol defines two port roles: ring protection link (RPL) owner port and common port. The link where an RPL owner port resides is called a ring

 Copyright © Huawei Technologies Co., Ltd.

protection link. The device where an RPL owner port resides is called an RPL owner node.

- **RPL owner port**

  An ERPS ring has only one RPL owner port, which can be configured by users. Loop avoidance is achieved on an ERPS ring by blocking the RPL owner port. A blocked RPL owner port cannot forward data but can send ring automatic protection switching (R-APS) messages.

  As defined in the IEEE 802.1ag protocol, the down maintenance end point (MEP) on the RPL owner port can receive and send Ethernet OAM messages.

  When an RPL owner node receives a fault message and learns that a node or link on the ERPS ring is faulty, the RPL owner node automatically enables the RPL owner port. Then, the RPL owner port resumes traffic reception and transmission, ensuring uninterrupted traffic.

  NOTE

  For load balancing purposes, it is recommended to select a device with the following characteristics as an RPL owner node: the device is located close to the user side; an equivalent (or approximately so) number of nodes are located on the link between each of the two ring ports of the device and the convergence node.

- **Common port**

  On an ERPS ring, all ports except the RPL owner port are common ports. Common ports are responsible for monitoring the status of their directly connected links, and notifying any link status changes to the ports on other nodes.

Port States

On an ERPS ring, a port has two states:

- **Forwarding**

  In forwarding state, a port can forward data, and receive and send R-APS messages.

- **Discarding**

  In discarding state, a port cannot forward data, but can receive and send R-APS, EFM, CFM, LLDP, and 802.1x messages.

R-APS Messages

Defined in the ITU-T G.8032 protocol, R-APS messages are protocol messages that run on ERPS rings to notify nodes of fault occurrence or fault clearing. R-APS messages include the following types:

- **R-APS signal fail (SF)**

  indicates a link failure. When detecting a link failure, a node sends an R-APS (SF) message to notify other nodes on the ring. When receiving an R-APS (SF), the RPL owner node unblocks the RPL owner port so that the RPL owner port forwards data. Other nodes determine whether to flush the FDB according to the DNF flag in the R-APS (SF) message.

- **R-APS no request (NR)**

  indicates a link recovery. When detecting a link recovery, a node sends an R-APS (NR) message to notify other nodes on the ring. When receiving an R-

 Copyright © Huawei Technologies Co., Ltd.

APS (NR), the RPL owner node starts the WTR timer. Other nodes determine whether to flush the FDB according to the DNF flag in the R-APS (NR) message.

- **R-APS no request, RPL blocked (NR, RB)**

  An R-APS (NR, RB) message is a type of R-APS (NR). It is sent by an RPL owner to signal that the network is normal and the RPL has been blocked. When receiving an R-APS (NR, RB), other nodes on the ring will unblock the common ports, and determine whether to flush the FDB according to the DNF flag in the message.

The following explains the acronyms involved in the description above.

Acronym	Full Name	Meaning
NR	No request	The network is in normal conditions.
RB	RPL blocked	An RPL is blocked.
DNF	Do not flush	A flag in the R-APS message fields and indicates whether to flush the FDB.
FDB	Filtering database	A forwarding database on a node.
SF	Signal fail	The network encounters a fault.

Timers

The ITU-T G.8032 protocol defines the following timers for an ERPS ring.

- **Guard timer**

  After a node on a Layer 2 network running ERPS recovers from a link fault or node fault, the node sends an R-APS (NR) message to other nodes on the ring and starts the guard timer. The node does not receive R-APS messages before the timer expires, which prevents the node from unblocking a recovered port after receiving expired R-APS (SF) messages. The unblocking will cause a loop on the network. If the node receives R-APS (SF) messages after the timer expires, the node changes its port state to forwarding.

- **Holdoff timer**

  Layer 2 networks running ERPS require different protection switching sequences. For example, when a node becomes faulty in a multiple-layer service application, carriers prefer a duration to rectify the fault and the duration is not perceived by subscribers. To meet the carrier requirement, protection switching is not performed immediately after a fault occurs. The fault is reported only if it is not rectified after the holdoff timer expires.

- **WTR timer**

  On a Layer 2 network running ERPS, an RPL owner port enters the forwarding state when a node or link on the ERPS ring fails. When the node or link recovers, some ports on the node or link may still stay in the Down state (that is, not Up yet). If the RPL owner port is immediately blocked upon the fault recovery, and finds that there are still ports in the Down state, the RPL owner port will enter the forwarding state again. To avoid such toggling protection

 Copyright © Huawei Technologies Co., Ltd.

states in the case of intermittent faults, the RPL owner port starts a wait-to-restore (WTR) timer when receiving the R-APS (NR) message from the other ports.

- If the RPL owner port receives the R-APS (SF) message from the other ports before the WTR timer times out, the RPL owner port stops the WTR timer and stays in the forwarding state.
- If the RPL owner port does not receive the R-APS (SF) message when the WTR timer times out, the RPL owner port is blocked and sends the R-APS (NR, RB) message to other ports. When receiving the R-APS (NR, RB) message, the other ports switch to the forwarding state.

16.8.3 Principle of ERPS

This topic uses ERPS network diagrams to describe the fault detection and recovery processes of an ERPS ring.

Network Diagram

Figure 16-40 shows the network diagram of an ERPS single ring. The nodes on the ERPS ring exchange R-APS messages to notify each other of the port status changes and coordinate to implement protection switching for services. The port status changes on a ring node depend on detection of local port status, Ethernet OAM detection results, and the port status changes on other nodes of the ring.

Figure 16-40 Network diagram of an ERPS single ring

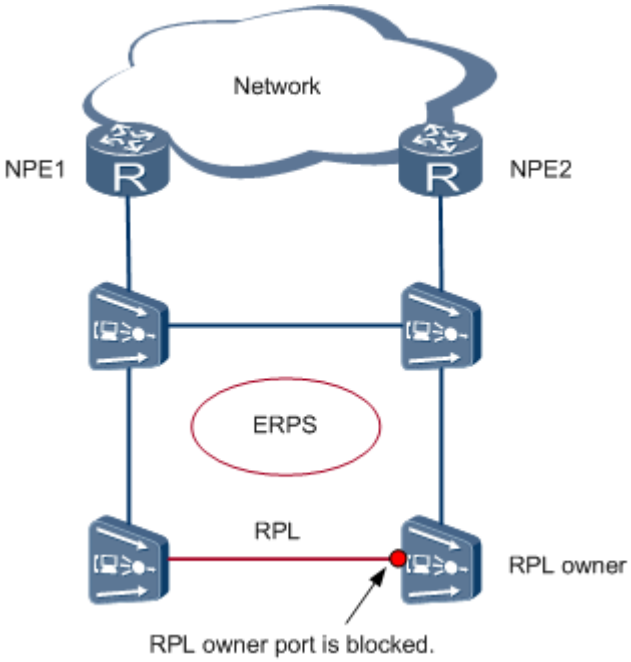
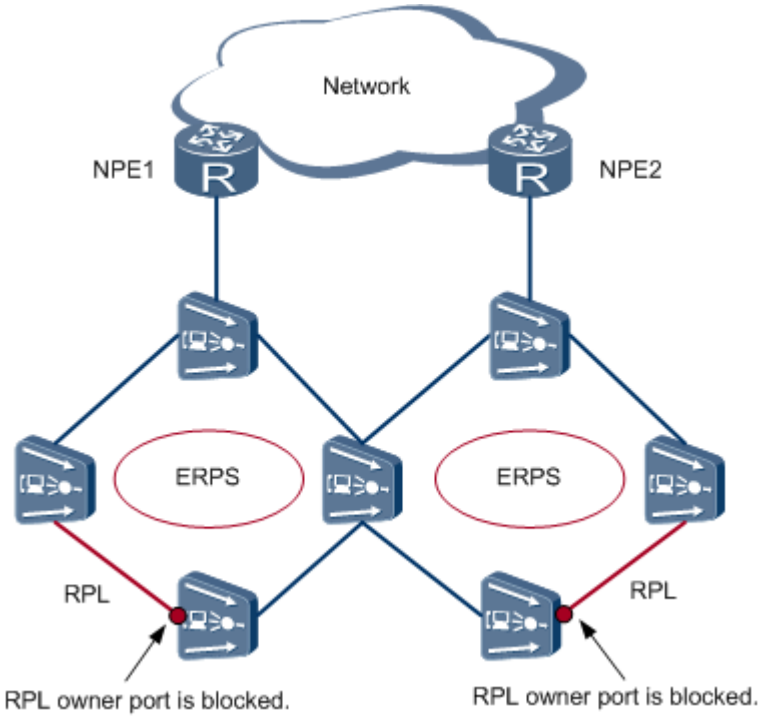


Figure 16-41 shows the network diagram of ERPS interconnected rings. The two physical ring networks are interconnected through one node. Four ports on the interconnection node are configured with ERPS. Two of the four ports locate on one ring, and the other two on the other ring. Except the interconnection node, other nodes on the two interconnected rings function in the same mechanism as the nodes on a single ring. The R-APS protocol messages of the two rings are independent of each other, so the two rings need different control VLANs.

 Copyright © Huawei Technologies Co., Ltd.

The interconnection node can be configured with multiple ERPS rings, and the rings can have different numbers of nodes.

Figure 16-41 ERPS rings interconnected through one node



Configuration Items of ERPS Ring

Each ring node must be configured with the following items:

- Control VLAN

  The control VLAN transmits ERPS messages and does not forward data messages, which improves ERPS security.
- Protection instance

  Protection instances are implemented by configuring mappings between MSTP instances and VLANs. On a Layer 2 network with ERPS enabled, the VLAN that transmits ERPS messages and data messages must be mapped to a protection instance so the ERPS ring forwards or blocks the messages based on the blocking principles. Without proper forwarding or blocking, the messages may cause a broadcast storm on the ring network and render the network unavailable.
- Port role

  Configuration related to port roles is to add ports to an ERPS ring, and specify the RPL owner port on the RPL owner node.

Fault Detection and Recovery

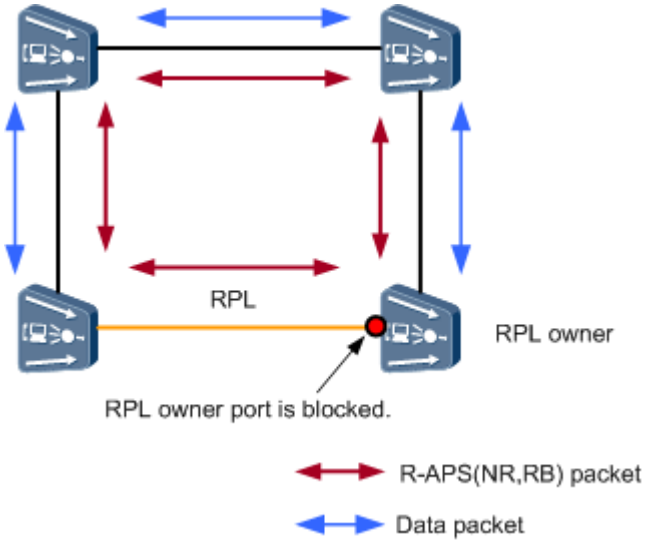
The following uses a single ERPS ring as an example to describe the fault detection and recovery processes of an ERPS ring.

In normal conditions, all links on the network are in the UP state, and the RPL owner port is blocked, therefore forming link redundancy backup and avoiding

 Copyright © Huawei Technologies Co., Ltd.

loops. Service traffic is forwarded through all links except the RPL. In this situation, the R-APS messages travelling on the network are all R-APS (NR, RB) messages that are sent by the RPL owner node, as shown in **Figure 16-42**.

Figure 16-42 An ERPS ring network in the normal state

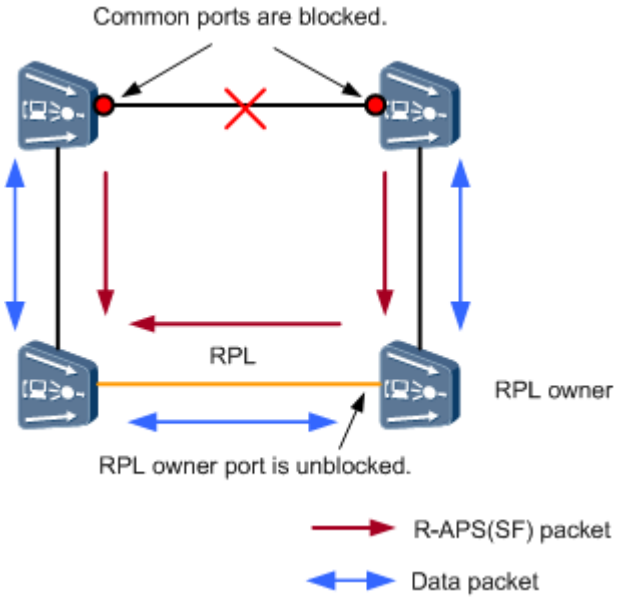


When a fault occurs on a link, the nodes connected to the faulty link detect the fault and start the holdoff timer. If the fault persists after the holdoff timer times out, the nodes block the ports on the faulty link, and send R-APS (SF) messages to other nodes on the ring. After receiving the R-APS (SF) message, the RPL owner node unblocks the RPL owner port, and switches to the RPL to transmit service traffic. In this way, ring network protection is implemented. When receiving the R-APS (SF) message for the first time, the node will flush its FDB, and the ring network enters the protection state, as shown in **Figure 16-43**. The node will not trigger any operation when receiving subsequent R-APS (SF) messages.

To implement fast switching, each node detecting the link fault will rapidly send 3 R-APS (SF) messages, at a 3.33 ms interval, which equals the interval of fast detection by using CCM messages. After sending the 3 R-APS (SF) messages, the node will send one R-APS (SF) message every 5s. Such a mechanism ensures fast network switching without burdening the devices.

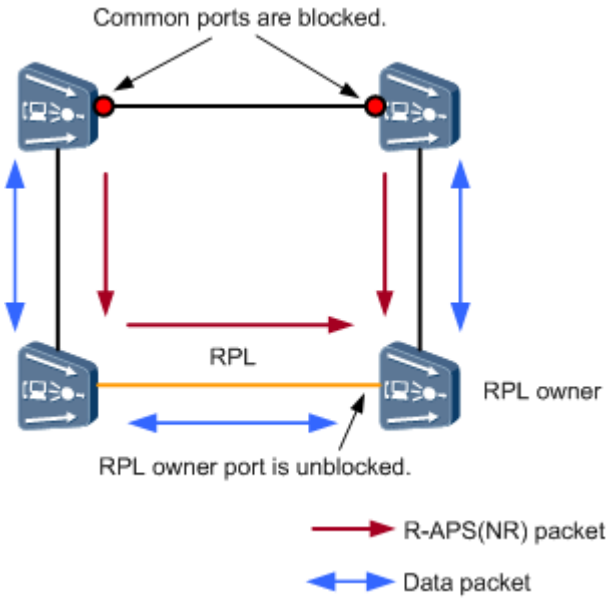
 Copyright © Huawei Technologies Co., Ltd.

Figure 16-43 An ERPS ring network in the protection state



When the fault recovers on the link, the nodes connected to the originally faulty link detect the fault recovery, start the guard timer, and stay in the blocked state. At the same time, the nodes send R-APS (NR) messages to the RPL owner node, notifying the RPL owner node that the RPL owner port can be blocked, as shown in **Figure 16-44**.

Figure 16-44 ERPS ring network status after fault recovers and before WTR timer times out



After receiving the R-APS (NR) message, the RPL owner node blocks the RPL owner port only after the WTR timer expires. Meanwhile, the RPL owner node sends the R-APS (NR, RB) message. The other nodes on the ring flush their FDB when receiving the R-APS (NR, RB) message. After the guard timer times out, the node that detects the fault recovery stops periodically sending R-APS (NR)

 Copyright © Huawei Technologies Co., Ltd.

messages and unblocks the local ports. When the network resumes the normal state, the RPL owner node continues to send R-APS (NR, RB) messages.

16.8.4 Configuring ERPS

The device supports ERPS as defined in ITU-T G.8032. Interoperation is achieved if all devices participating in the ring network support ERPS. The ERPS feature enables fast convergence and meets carrier-class reliability requirements.

Prerequisites

The service VLANs to be protected have been configured.

Context

- In the single ERPS ring scenario, the ring must be configured on all nodes participating in the ring. In the multiple-ERPS-ring scenario, each ring has its own configurations.
- Note that the system does not have a restriction mechanism for some configuration items. When planning and configuring such items, make sure that they are planned and configured correctly, in line with system requirements. The following table provides the configuration restrictions on ERPS.

Table 16-11 ERPS configuration restrictions

Item	Limitation	Does the Device Check the Limitation?
Control VLAN	• A control VLAN must be a standard VLAN and cannot be used by a Rapid Ring Protection Protocol (RRPP) ring or a smart link. • A control VLAN cannot be a reserved VLAN. • An ERPS ring cannot be modified or deleted if a port is added to the ring. To modify or delete a control VLAN, delete the port from the ERPS ring first. • Different ERPS rings must be configured with control VLANs of different IDs on the same node.	Yes

 Copyright © Huawei Technologies Co., Ltd.

Item	Limitation	Does the Device Check the Limitation?
	• The control VLAN must be different from the service VLAN. • A control VLAN cannot be configured as a native VLAN. • The attribute of a control VLAN must be common. • All the devices in an ERPS ring must be configured with the same control VLAN.	No. The user must correctly configure this item.
Ring port	• Each node supports up to 2 ports to join the same ERPS ring. • MSTP, RRPP, smart link, Ethernet port protection, and ERPS are mutually exclusive. • Upstream ports and cascade ports can function as ring ports, but user-side ports cannot.	Yes

 Copyright © Huawei Technologies Co., Ltd.

Item	Limitation	Does the Device Check the Limitation?
	• When a control VLAN is created for an ERPS ring or a port is added to an ERPS ring, the port is not added to the control VLAN forcibly. Ensure that all the ports added to the ERPS ring are in the control VLAN. If a port is not in the control VLAN, the ERPS packets sent to the port are discarded. • A ring port must not be configured with static MAC addresses or static ARP entries. • On all nodes except the convergence node, it is not allowed to isolate two ring ports using the port isolation function. • On all nodes, ensure that packets of the control VLAN can be transmitted and received between two ring ports of the same node. • The ring ports on the interconnection node of an access-layer ERPS ring and a convergence network can be configured as cascade ports or upstream ports, but the ring ports on other nodes on the rings can be configured as upstream ports only. • In a network where an access-layer ERPS ring is connected to the upstream network through multiple nodes, the ring ports on all nodes of the ERPS ring must be upstream ports. • The interconnection node of an access-layer ERPS ring and a convergence network does not support associated routes or Neighbor Discovery (ND) entries. If the interconnection node is enabled to automatically generate associated routes and ND entries, when the topology of the ERPS ring changes, the DHCPv6 users under all nodes on the ring must dial up again. Otherwise, services may fail.	No. The user must correctly configure this item.

 Copyright © Huawei Technologies Co., Ltd.

Item	Limitation	Does the Device Check the Limitation?
Protection instance	• One ERPS ring can be configured with one or multiple protection instances. Each protection instance maps one or multiple VLAN services. • One ERPS ring can be configured with one or multiple protection instances. Each protection instance maps one or multiple VLAN services.	No. The user must correctly configure this item.

Procedure

Step 1 Run the **vlan** command to create a control VLAN.

Step 2 Run the **port vlan** command to add ERPS ring ports to the control VLAN.

Step 3 Map the control VLAN and service VLANs to the protection instances.

1. Run the **stp region-configuration** command to enter the MSTP region mode.
2. Run the **instance vlan** command to map the service VLANs and the control VLAN to the specified MSTP instance.
3. Run the **active region-configuration** command to activate the configuration of the MSTP region.
4. Run the **quit** command to quit the MSTP region mode.

Step 4 Run the **erps ring** to create an ERPS ring and enter the ring mode.

The nodes on the same ring must be configured with the same ring ID.

Step 5 Run the **control-vlan** command to configure the control VLAN for the ERPS ring.

vlan-id must be the same as the *vlan-id* specified in the **vlan** command.

Step 6 Run the **protected-instance** command to configure the protection instance for the ERPS ring.

instance-id must be the same as the *instance-id* specified in the **instance vlan** command.

Step 7 Run the **port** command to add the ports to the ERPS ring.

port-id must be the same as the *port-id* specified in the **port vlan** command. Specify the **rpl owner** parameter when configuring the RPL owner port.

Step 8 (Optional) Configure ERPS parameters.

NOTE

You are advised to configure the same ERPS parameters on all nodes of the ring.

- Configure the timers.
  - Run the **guard-timer** command to configure the guard timer started by a node when the node detects a fault recovery.

 Copyright © Huawei Technologies Co., Ltd.

  - Run the **holdoff-timer** command to configure the holdoff timer started by a node when the node detects a fault.
  - Run the **wtr-timer** command to configure the WTR timer started by a node when the node receives an R-APS (NR) message.
- Run the **description** command to configure the description of the ERPS ring.
- Run the **priority** command to set the 802.1p priority of ERPS messages on the ERPS ring.
- Run the **raps-mel** command to set the value for the MEL field in the ring automatic protection switching (RAPS) messages on the ERPS ring.

Step 9 Run the **display erps** command to query the ERPS configuration.

----End

Example

The following is an example of configuring the RPL owner node by using the following ERPS parameters:

- ERPS ring ID: 1
- Control VLAN ID: 20
- Protection instance ID: 3; mapping service VLAN IDs: IDs 2-10, 12-16
- Ring ports: 0/9/0 and 0/9/1, of which 0/9/0 is the RPL owner port
- ERPS ring description: ERPS Ring 1. The other parameters use their default values.

```
huawei(config)#vlan 20 standard
huawei(config)#port vlan 20 0/9 0,1
huawei(config)#stp region-configuration
huawei(stp-region-configuration)#instance 3 vlan 2 to 10 12 to 16 20
huawei(stp-region-configuration)#active region-configuration
huawei(stp-region-configuration)#quit
huawei(config)#erps ring 1
huawei(config-erps-ring1)#control-vlan 20
huawei(config-erps-ring1)#protected-instance 3
huawei(config-erps-ring1)#port 0/9/0 rpl owner
huawei(config-erps-ring1)#port 0/9/1
huawei(config-erps-ring1)#description ERPS Ring 1
huawei(config-erps-ring1)#quit
huawei(config)#display erps
  --------------------------------------------------------------------------
  Ring  Control  WTR Timer  Guard Timer  Port 1          Port 2
  ID    VLAN     (min)      (csec)
  --------------------------------------------------------------------------
    1       20         5         200  (D,R)0/9/0      (D)0/9/1
  --------------------------------------------------------------------------
  D : Discarding, F : Forwarding, R : RPL Owner
  Total number of rings configured : 1
```

16.8.5 ERPS Reference Standards and Protocols

The reference standards and protocols of the ERPS feature are as follows:

- ITU-T G.8032 V1
- ITU-T G.8032 V2

 Copyright © Huawei Technologies Co., Ltd.

16.9 VRRP Snooping

VRRP is a fault-tolerant protocol. It allows multiple routers to form a virtual routing device, and provides a mechanism, which ensures that services will be taken over in time by another device once the next hop of a host fails. In this way the continuity and reliability of communication are ensured. VRRP snooping is to snoop (or listen for) VRRP packets.

16.9.1 Introduction to VRRP Snooping

Definition

Virtual Router Redundancy Protocol (VRRP) is a fault-tolerant protocol. It allows multiple routers to form a virtual routing device, and provides a mechanism, which ensures that services will be taken over in time by another device once the next hop of a host fails. In this way the continuity and reliability of communication are ensured.

VRRP snooping is to snoop (or listen for) VRRP packets. According to VRRP packets the listening device can confirm the port to which the upstream master router is connected. Then, the listening device will transmit the unicast service stream to the master router and at the same time transparently transmit the VRRP packets of any of other routers to another router in the same VRRP group.

Purpose

To enhance system reliability, the MA5800 is directly dual-homed to two or more BRASs in the upstream direction, and the BRASs run the VRRP protocol. When the MA5800 works in the SVLAN+CVLAN forwarding mode and MAC address learning is disabled, the upstream ports of the MA5800 need to be isolated from each other in order to prevent unknown unicast broadcast storm. However, when the upstream ports are isolated, the upstream BRASs cannot interoperate VRRP packets. VRRP snooping is adopted for forwarding VRRP packets because VRRP snooping enables the BRASs to interoperate VRRP packets so that the BRASs can run VRRP normally.

To prevent asynchronous VRRP status, Huawei develops the VRRP Group Management Protocol (VGMP), which is extended based on VRRP. VGMP manages the VRRP status of each backup group in a unified manner. VGMP provides a mechanism for managing the status synchronization, preemption, and channels of multiple VRRP backup groups. When the MA5800 works in the SVLAN+CVLAN mode, the VGMP protocol may fail to run if the upstream ports are isolated. To enable the BRASs to interoperate VGMP packets, the MA5800 can transparently transmit VGMP packets.

16.9.2 Principle of VRRP Snooping

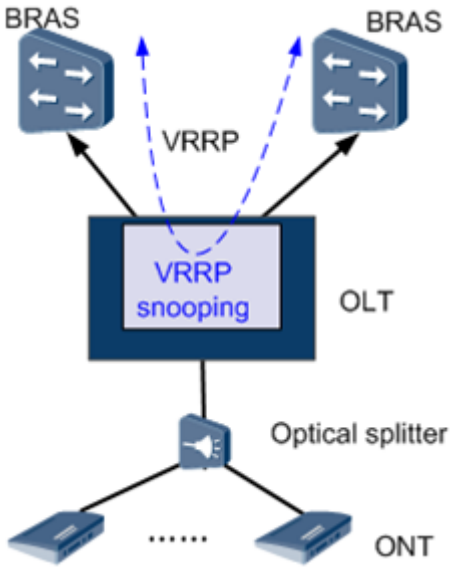
Basic Principle of VRRP Snooping

In a network, the failure of a single router may cause failure of the entire network. To address this problem, multiple VRRP-running routers on the upper layer of the

 Copyright © Huawei Technologies Co., Ltd.

MA5800 can form a set of virtual routers. Thus, another router will automatically take over the forwarding service if the master router fails. Viewed from the MA5800 there is still only one router, but this router may be interconnected to two upstream ports of the MA5800. When the path switches from one port to the other, the upstream router will send free ARP packets to update the forwarding path.

Figure 16-45 Network topology of VRRP snooping



As shown in **Figure 16-45**, to solve the problem of MAC address insufficiency, the MA5800 adopts the SVLAN+CVLAN forwarding mode. In this forwarding mode, MAC address learning needs to be disabled, and the upstream ports of the MA5800 need to be isolated to avoid broadcast storm of unknown unicast packets.

After the upstream ports of the MA5800 are isolated, the multiple upstream routers directly connected to the MA5800 cannot forward VRRP packets to each other through the upstream port of the MA5800. The result will be that the routers fail to run the VRRP protocol. To address this problem, the MA5800 needs to use software forwarding in order to implement VRRP protocol packet exchange between the isolated ports.

When MAC address learning is disabled, packets going upstream may be forwarded to the two upstream ports at the same time, which is a waste of bandwidth. In this case, a static MAC address needs to be configured so that unicast packets are forwarded to the master router only. The MA5800 listens to VRRP packets and free ARP packets to learn the upstream port to which the master router is currently connected. By using the static MAC address, the MA5800 forwards Layer 2 service data to this upstream port.

When the router sends free ARP packets to the MA5800 for switching the forwarding path, the free ARP packets may be lost due to network reasons. If ARP packets are lost, the MA5800 listens to VRRP packets to update the ARP entry. This prevents a condition where Layer 3 forwarding services are interrupted for a long time because the forwarding path is not updated in time.

 Copyright © Huawei Technologies Co., Ltd.

VRRP Snooping in the VLAN+MAC Forwarding Mode

In the VLAN+MAC forwarding mode, the two routers to which the MA5800 is dual-homed can run the VRRP protocol without additional processing on the MA5800 as long as the following condition is met: The two upstream ports connected to the two routers can interoperate, which allows for normal forwarding of VRRP packets between the two routers. When the network condition is good, or when the MA5800 needs not consider the loss of free ARP packets (a router may provide for retransmission of free ARP packets), VRRP snooping needs not be enabled on the MA5800 when the MA5800 runs in the VLAN+MAC forwarding mode.

16.9.3 Configuring VRRP Transparent Transmission in the S+C Forwarding Mode

In the S+C forwarding mode, after VRRP snooping is enabled, VRRP packets can be forwarded between two isolated upstream ports.

Context

To enhance the system reliability, the MA5800 is directly connected to two or more routers in the upstream direction for dual homing.

- In the S+C forwarding mode, the MAC address learning function needs to be disabled. In addition, to prevent broadcast storms, two upstream ports of the MA5800 need to be isolated. Therefore, the VRRP packets between upper-layer routers cannot be forwarded through the upstream ports of the MA5800. To solve this problem, enable the VRRP transparent transmission function.
- In the VLAN+MAC forwarding mode, the VRRP transparent transmission function may not be enabled.

Procedure

Step 1 Configuring an isolation group.

Run the **isolate group** command to configure an isolation group. To avoid forwarding downstream service packets to another upstream port, you can add the upstream ports to an isolation group.

Step 2 Configure a snooping port.

Run the **vrrp-snoop port** command to configure the upstream port connecting the MA5800 and the router as a snooping port.

Step 3 Configure the virtual IP address and VLAN to be snooped.

Run the **vrrp-snoop ip** command to configure the IP address and VLAN of the virtual router to be snooped.

Step 4 Enable VRRP snooping.

Run the **vrrp-snoop enable** command to enable VRRP snooping.

----End

 Copyright © Huawei Technologies Co., Ltd.

Example

Assume the following configurations: The MA5800 is connected to two routers through upstream ports 0/9/0 and 0/9/1. The VRRP packets between the routers need to be transparently transmitted through the upstream ports of the MA5800. The VLAN forwarding mode is S+C, the VLAN ID is 100, and the IP address of the virtual router is 10.71.10.1. To perform these configurations, do as follows:

```
huawei(config)#isolate group port 0/9/0 0/9/1
huawei(config)#vrrp-snoop port 0/9/0
huawei(config)#vrrp-snoop port 0/9/1
huawei(config)#vrrp-snoop ip 10.71.10.1 vlan 100
huawei(config)#vrrp-snoop enable
```

16.9.4 VRRP Snooping Reference Standards and Protocols

The reference standards and protocols of this feature are as follows:

- RFC3768, Virtual Router Redundancy Protocol (VRRP)
- RFC2787, Definitions of Managed Objects for the Virtual Router Redundancy Protocol

16.10 Ring Check

The ring check feature is mainly used to detect the user-side, cascade-side, and network-side ring networks, and eliminate the user-side ring network.

16.10.1 Introduction to Ring Check

Definition

The Ring Check feature enables the device to send the Ring Check packets periodically and monitors the Ring Check packets received on the user-side, cascade-side, and network-side to detect whether a loop occurs on the network of the carrier. If a loop occurs, the MA5800 deactivates the subscriber port on the loop and reports the corresponding alarm to the NMS. This ensures that the device runs in the normal state and that legal users are not affected.

Purpose

Ring check is used to quickly locate the user-side, cascade-side, and network-side ring network, and eliminate the ring network according to requirements.

- To prevent the self-loop on a single user port from occurring
- To prevent the loop between user ports from occurring
- To prevent the loop between a user port and a network port, or between a subscriber port and a cascade port from occurring

Benefit

For carriers

The Ring check feature enables the system to detect the carrier's network and report an alarm to the NMS when a loop occurs. The alarm enables the carrier to

 Copyright © Huawei Technologies Co., Ltd.

know the network fault in the shortest period of time so that the fault can be quickly rectified to resume the normal running of the network.

For users

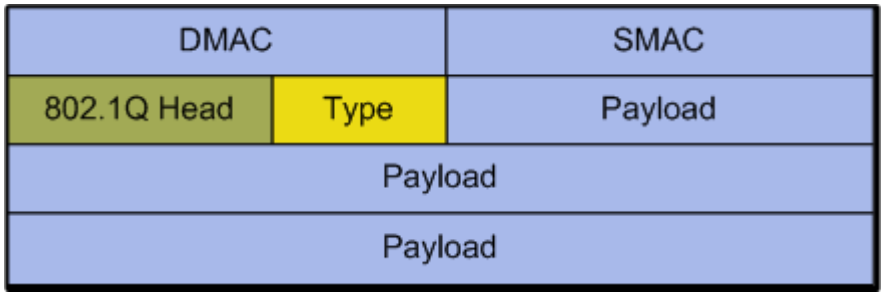
The Ring check feature enables the device to deactivate the user port on a loop to ensure that authorized users receive a good network service rather than be affected.

16.10.2 Principle of Ring Check

Format of the Ring Check Packet

The ring check feature enables the device to transmit the ring check packets to the user port, cascade side, and network side port periodically, and to monitor the ring check packets received on the user side, cascade side, and the network side to check whether a loop occurs on the network of the carrier. **Figure 16-46** shows the format of the ring check packet.

Figure 16-46 Format of the ring check packet



- DMAC indicates the broadcast MAC address with value ff-ff-ff-ff-ff-ff and SMAC indicates the bridge MAC address.
- 802.1Q Head is optional according to flow attributes on the user side.
- Type indicates the proprietary Ethernet type, which can be configured.
- Payload of the packet content is proprietary and it needs not be configured.

Principle

After the ring check function is enabled on the user side, cascade side, and network side, the device periodically transmits private ring check packets to the user port, the cascade port, and the network port, and captures the user-side, cascade-side, and network-side ring check packets on the network and user sides simultaneously.

- For the ring check packets (that are sent from the user side, cascade-side, or network side of the local device) captured on the cascade side or network side, the system reports an alarm to the NMS.
- For the user-side, cascade-side, and network-side ring check packets captured on the user side, the system reports an alarm to the NMS and deactivates the port receiving the packets, eliminating the ring network.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

After the fault is solved, the port will be activated after some time. If you want to enable the port quickly, the port needs to be deactivated and activated again.

- Detects a maximum of 12 up traffic streams per second. If the system has 8000 up traffic steams, 666.67s (8000/12) is required for detecting a loop if a loop exists.

Figure 16-47 shows the use-side ring network scenarios in FTTH/DSLAM applications.

- In the case of (1), (2), and (3), after receiving the ring check packets captured on the user side, the system reports an alarm to the NMS and deactivates the port receiving the packets, eliminating the ring network.
- In the case of (4), after detecting the ring network between the network side or cascade side and user side, the system reports an alarm to the NMS and deactivate the port on the user side, eliminating the ring network.

Figure 16-47 Use-side ring network scenarios in FTTH/DSLAM applications

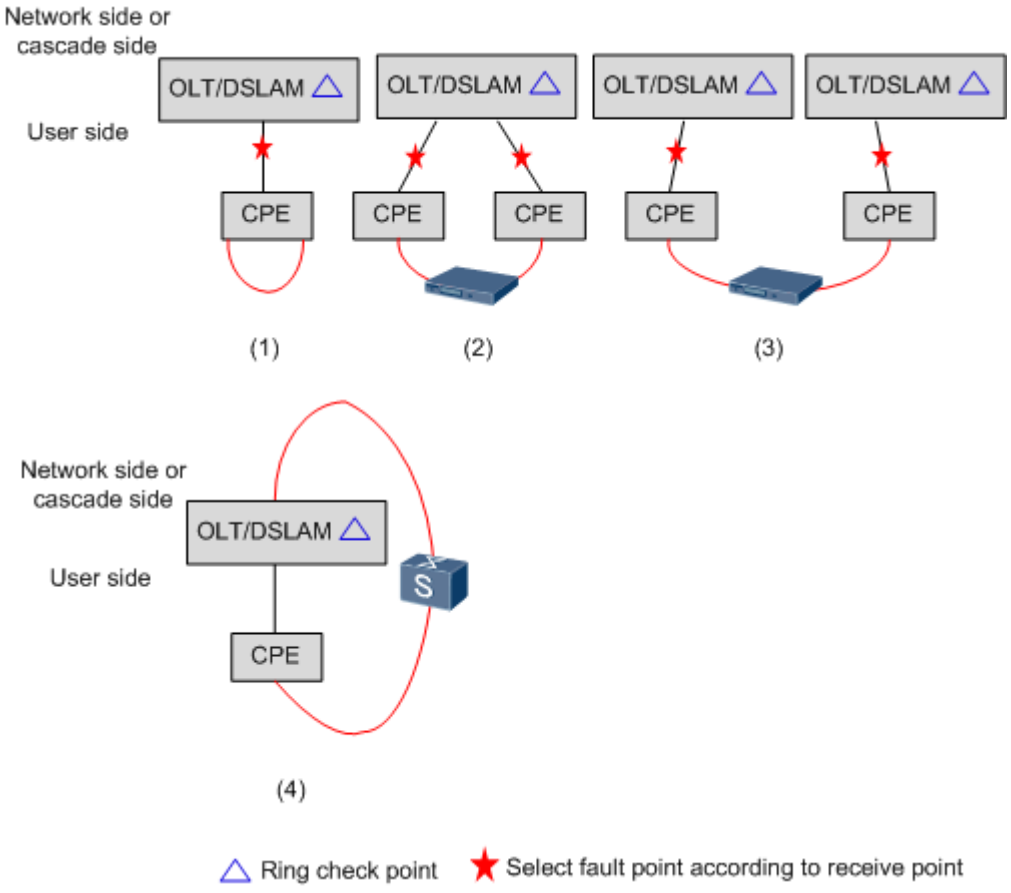


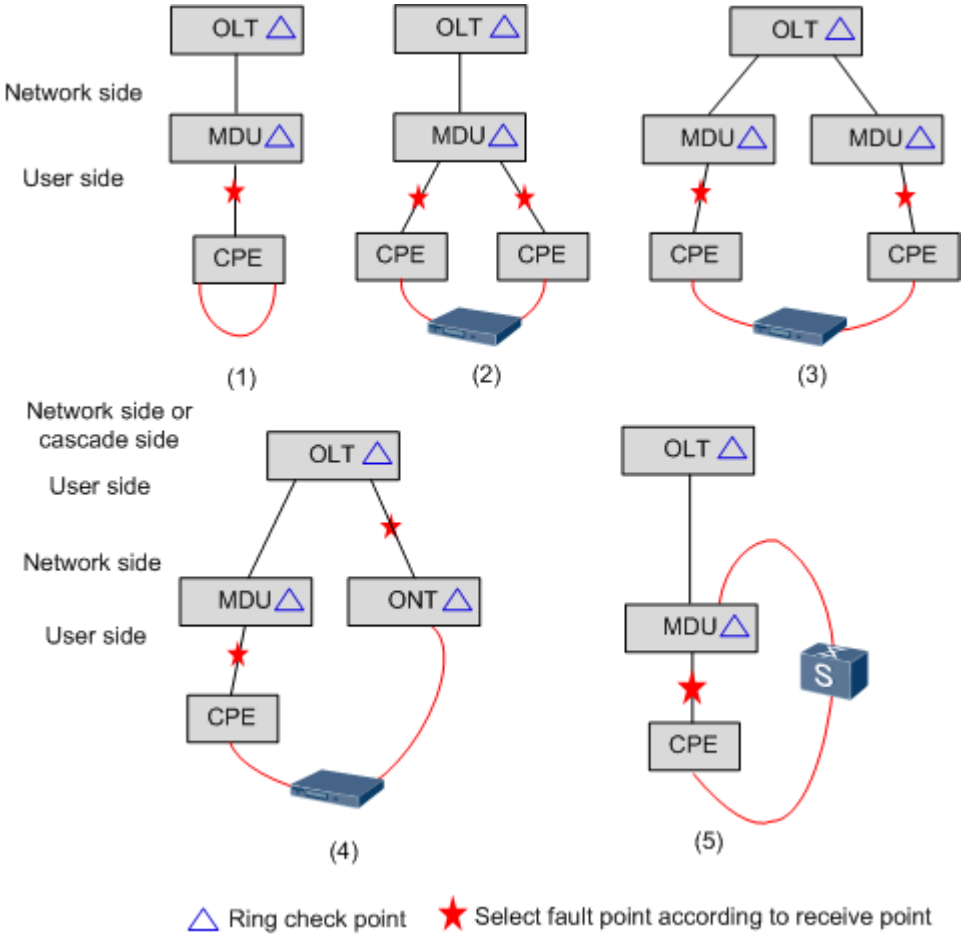
Figure 16-48 shows the use-side ring network scenarios in FTTB/FTTC applications.

- The Ethernet Types of ring check packets of the OLT and the MDU are recommended to be the same.

 Copyright © Huawei Technologies Co., Ltd.

  - If Ethernet Types of ring check packets of the OLT and the MDU are the same, the ring check packets are captured on the OLT and the MDU and are judged.
  - If Ethernet Types of ring check packets of the OLT and the MDU are different, the OLT and the MDU capture its own ring check packets. In the case of (4), the system cannot detect a ring network because the MxU and OLT do not identify ring check packets sent by the peer.
- In the case of (1), (2), (3), and (4), after receiving the ring check packets captured on the user side, the system reports an alarm to the NMS and the system deactivates the port receiving the packets, eliminating the ring network.
- In the case of (5), the MDU captures ring check packets sent from its user side on the network side, reports an alarm to the NMS, and sends network-side ring check packets to the user side. After receiving ring check packets sent from the network side, the user side deactivates the port receiving the packets, eliminating the ring network.

Figure 16-48 Use-side ring network scenarios in FTTB/FTTC applications



 Copyright © Huawei Technologies Co., Ltd.

16.10.3 Configuring the Ring Network Detection

This topic describes how to configure the ring network detection on the user side, cascade-side, and network side to prevent the services from being affected by the ring network.

Context

- By default, the ring network detection is disabled.
- After the ring network detection is enabled, the system automatically detects the ring network.

NOTICE

To ensure the security of the device, it is recommended that the ring network detection is enabled.

Procedure

Step 1 (Option) Run the **ring check private-ethtype** command to configure the private Ethernet protocol type of ring network detection packets. If the default Ethernet protocol type (0x8300) of ring network detection packets is the same as the existing protocol type on the network, change the protocol type of the ring detection packets to be different from the protocol type on the network.

Step 2 (Option) Run the **ring check resume-interval** command to set the auto-activation interval for the port in ring check.

After the auto-activation interval is set, when detecting a ring network, the system automatically activates the port that sends the ring check packet.

Step 3 (Option) Run the **ring check trap-notify** command to set the user-side ring check trap notification mode to alarm.

NOTE

By default, the user-side ring check trap is notified as an event.

Step 4 Enable the ring network detection.

- Run the **ring check enable** command to enable the user-side ring network detection.
- Run the **ring check uplink detection enable** command to enable the network-side ring network detection.
- Run the **ring check cascade detection enable** command to enable the cascade-side ring network detection.

Step 5 Run the **display ring check config** command to query the status of the ring network detection.

Step 6 Run the **display ring check record** command to query the records of the ports, ONTs, or CNUs which were deactivated after ring check is enabled.

----End

 Copyright © Huawei Technologies Co., Ltd.

Example

To enable the ring network detection, do as follows:

- Set the private Ethernet type of the ring check packet to 0x2233.
- Set the auto-activation interval to 200 minutes.
- Set the user-side ring check trap notification mode to alarm.

```
huawei(config)#ring check private-ethtype 0x2233
huawei(config)#ring check resume-interval 200
huawei(config)#ring check trap-notify alarm
huawei(config)#ring check enable
huawei(config)#ring check uplink detection enable
  Warning: If the interconnected port is in user role and user-side ring check
is enabled, this operation may lead to shutdown of the interconnected port. Are
you sure you want to continue?(y/n)[n]:y
huawei(config)#ring check cascade detection enablering check cascade detection enable
  Warning: If the interconnected port is in user role and user-side ring check
is enabled, this operation may lead to shutdown of the interconnected port. Are
you sure you want to continue?(y/n)[n]:y
huawei(config)#display ring check config
  Ring check switch status        : enable
  Ring check private ethtype      : 0x2233
  Ring check resume-interval(min) : 200
  Ring check trap-notify          : alarm
  Ring check uplink detection     : enable
  Ring check cascade detection    : enable
```

16.11 Monitor Link

A monitor link group is composed of one uplink and several downlinks. It is used to monitor the fault of uplinks.

16.11.1 Introduction to Monitor Link

Definition

A monitor link group is composed of one uplink and several downlinks. It is used to monitor the fault of uplinks. After a monitor link group is created, the downstream device can defect the fault of upstream link and switch to a normal uplink.

Purpose

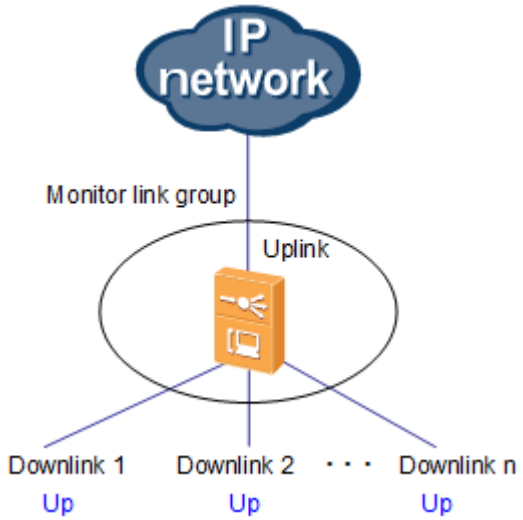
The monitor link group is used in the scenario when a device on an aggregation point needs notify uplink faults to cascaded devices. A downstream device working in dual uplink mode can switch to the backup uplink according to the fault status notified by the monitor link group.

 Copyright © Huawei Technologies Co., Ltd.

16.11.2 Principles of Monitor Link

Basic Concepts

Figure 16-49 Composition of a monitor link group



The following describes some basic concepts related to the monitor link feature.

- Monitor link group

  A monitor link group is composed of one uplink and several downlinks.

  NOTE

  The link in a monitor link group may not be a single link, but may be a certain type of link group. The uplink can be an aggregation group or protect group. The downlink can only be a single link. The status of the downlink changes according to the status of the uplink.

- Uplink

  When the uplink in a monitor link group fails, it indicates that the monitor link group fails. In this case, the downlinks in the monitor link group will be blocked by force.

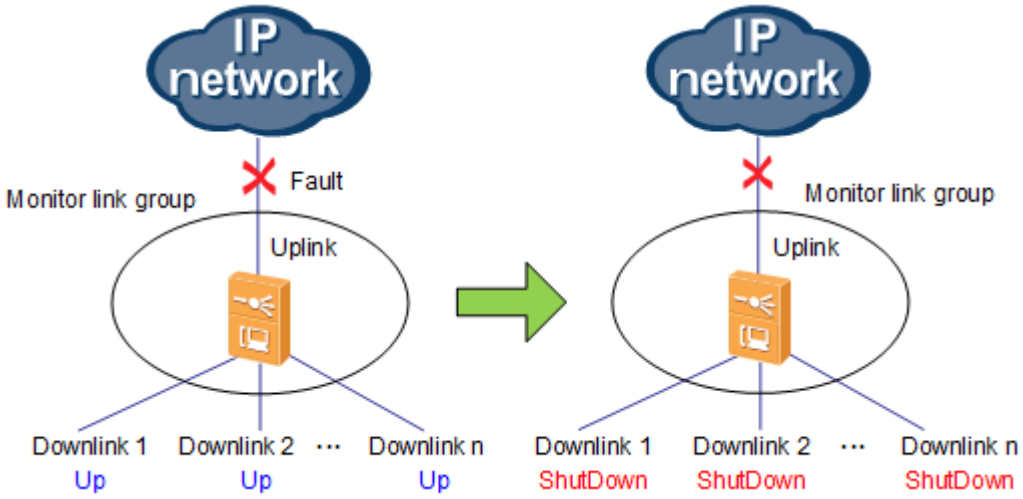
- Downlink

  When a downlink in a monitor link group fails, it does not affect the uplink or the other downlinks.

 Copyright © Huawei Technologies Co., Ltd.

Working Principle

Figure 16-50 Working principle of the monitor link feature



After a monitor link group is configured, its uplink will be monitored in real time. Once the uplink fails, all the UP downlinks in the monitor link group will be blocked by force. When the uplink recovers from the fault, the downlinks are resumed.

When the uplink is an aggregation group or protect group, the uplink is considered failed only when the entire aggregation group or protect group fails.

16.11.3 Configuring the Monitor Link Group

Procedure

Step 1 Run the **monitor-link group** command to create a monitor link group, and enter the monitor link group mode.

Step 2 Run the **member port** command to add members to a monitor link group, including the uplink and downlinks.

Step 3 (Optional) Run the **recover-time** command to configure the recovery time of a monitor-link group. The recovery time of a monitor-link group refers to the jitter-preventing time after which the faulty up link of the monitor-link group becomes normal.

Step 4 Run the **display monitor-link group** command to query the information about the monitor link group.

----End

Example

Assume the following configurations: The monitor link group ID is 1. The uplink is 0/9/0. The downlink with ID 0 is 0/9/1. The downlink with ID 1 is the Ethernet port

 Copyright © Huawei Technologies Co., Ltd.

of ONT. The ONT ID is 0. The Ethernet port ID of ONT is 1. The recover time of monitor link is 10 seconds.

```
huawei(config)#monitor-link group 1
huawei(monitor-link-group-1)#member port 0/9/0 uplink
huawei(monitor-link-group-1)#member port 0/9/1 downlink 0
huawei(monitor-link-group-1)#member port 0/10/0 ont 0 eth-port 1 downlink 1
huawei(monitor-link-group-1)#recover-time 10
huawei(monitor-link-group-)#quit
huawei(config)#display monitor-link group 1
  ---------------------------------------------------------------------------
  Group ID            : 1
  Recover Timer(s)    : 10
  Group State         : Normal
  Downlink Number     : 1
  ---------------------------------------------------------------------------
  Member          Role          Link    Last-up-time       Last-down-time
  F/S/P   ONT Port
  ---------------------------------------------------------------------------
  0/9/0           Uplink        Online  2017-07-19         2017-07-19
                                        17:54:28+08:00     17:53:53+08:00
  0/10/0    0   1 Downlink(0)   Online  -                  -
  0/9/1           Downlink(0)   Online  2017-07-19         2017-07-19
                                        17:54:33+08:00     17:53:58+08:00
  ---------------------------------------------------------------------------
```

 Copyright © Huawei Technologies Co., Ltd.

17 802.1X Authentication

17.1 Why Is 802.1X Authentication Required?
17.2 Device Roles Defined in 802.1X
17.3 Port-based Authentication
17.4 802.1X Information Exchange
17.5 Configuring 802.1X Authentication

17.1 Why Is 802.1X Authentication Required?

With the popularization of networks, more and more enterprises, campuses, communities, and SOHO users select Ethernet, which is a cost-effective, convenient, and high-bandwidth access mode and one of the most important access modes on the current data communication networks. Users can access network resources from any computer that is connected to a network, which brings severe security risks to the network environment. In addition, the implementation and maintenance of different network services require the underlying Ethernet to provide a necessary security authentication mechanism. How to correctly process user access rights becomes an increasingly prominent problem.

802.1X is used to authenticate and authorize the users connected to a LAN to protect network security. IEEE 802.1X (802.1X for short) is an IEEE standard for port-based network access control (PNAC). It provides an authentication mechanism for user access networks. By authenticating the rights of a user connected to a port, 802.1X can determine which network resources are accessible to the user device connected to the port.

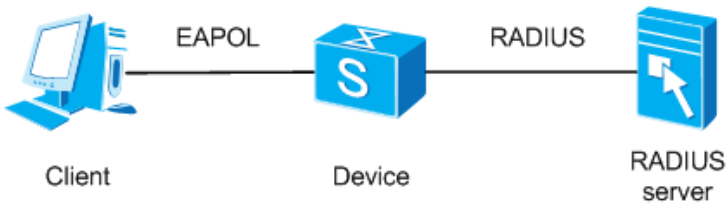
802.1X has complete user authentication and management functions, which can well support the accounting, security, operation, and management requirements of broadband networks.

17.2 Device Roles Defined in 802.1X

Figure 17-1 shows the roles of network devices that use the 802.1X authentication protocol.

 Copyright © Huawei Technologies Co., Ltd.

Figure 17-1 Device roles defined in 802.1X



- **Client:** A user terminal that seeks to gain access to the network. Based on whether supporting proactive initiation of 802.1X authentication, terminals are classified into common terminals and dumb terminals.
  a. Common terminals, such as office PCs, can run 802.1X authentication client software. Authentication is triggered after a terminal user enters its user name and password. If the authentication is successful, the terminal accesses the network.
  b. Dumb terminals, such as network printers and cameras, cannot run 802.1X authentication client software. They cannot access any network by proactively triggering authentication. Instead, they access networks through MAC address bypass (MAB) authentication.
- **Device:** It provides a port for the client to access the LAN and controls the client access rights or scope on the network based on the authentication status of the client. Located between the client and the authentication server, the device requests identity information from the client, encapsulates and decapsulates the Extensible Authentication Protocol (EAP) frame, interacts with the authentication server, and relays the response to the client.
- **Authentication server:** It authenticates the client and notifies the device whether the to-be-authenticated port may access the network, thereby implementing authentication, authorization, and accounting on the user. Generally, the authentication server is a RADIUS server.

After receiving Extensible Authentication Protocol over LAN (EAPOL) frames, the device strips off Ethernet headers and re-encapsulates the remaining EAP frames in RADIUS format (EAPOR) without modifying or checking the EAP frames. Then, the device relays the EAP frames to the authentication server, and the authentication server parses the EAP frames. In the reverse direction, when receiving the frames from the authentication server, the device removes the frame headers and leaves the EAP frames, encapsulates the frames into Ethernet frames, and sends them to the client.

17.3 Port-based Authentication

When 802.1X is enabled on a port but authentication is not triggered, all services on the port go through the guest VLAN. For terminals that are not 802.1X-capable, the guest VLAN is used to download 802.1X dial-up software or access restricted web pages. After 802.1 authentication is triggered, the system authenticates the client as follows:

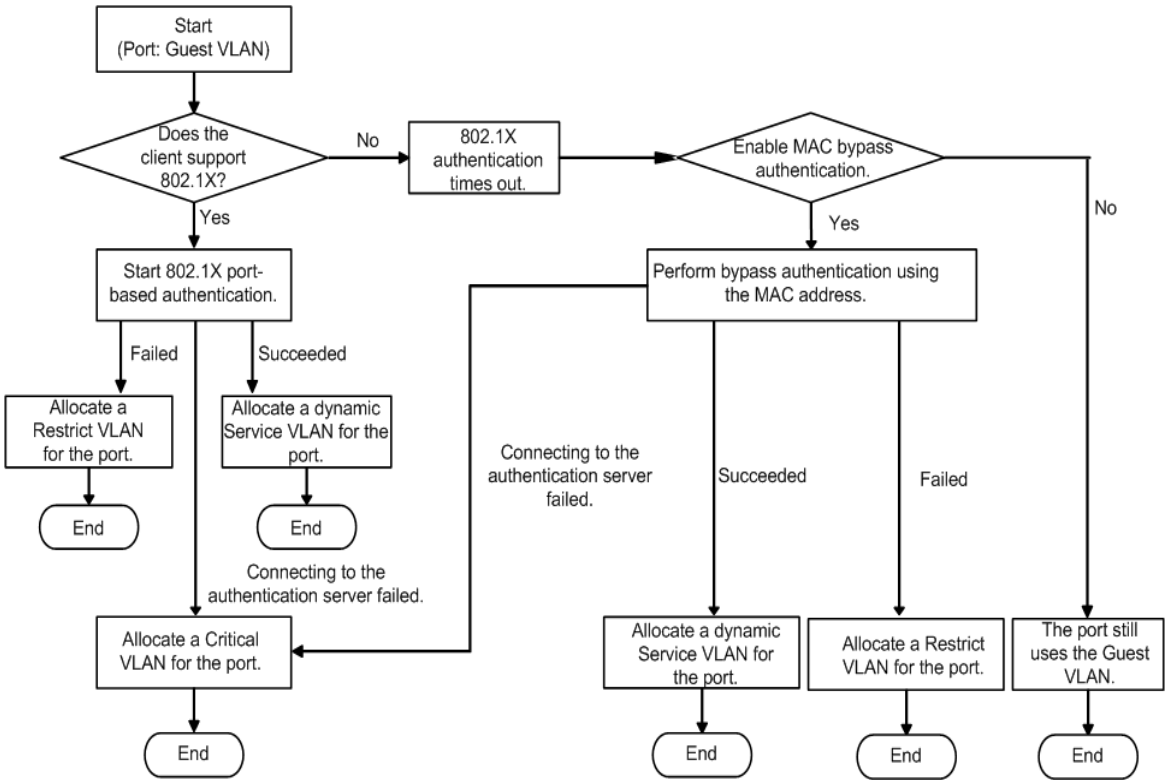
- If the client supports 802.1X client software and the client identity is legitimate, the 802.1X authentication is successful. Then, the client port is assigned a dynamic service VLAN, and the client can access the network.

 Copyright © Huawei Technologies Co., Ltd.

- If 802.1x authentication times out during EAPOL message exchange and MAC bypass (MAB) authentication is enabled, the device relays the MAC address of the client to the authentication server for authorization. If the MAC address of the client is legitimate, the authorization is successful and a dynamic service VLAN is assigned to the client. Then, the client can access the network.
- If 802.1X authentication of the port times out and the MAC address authentication is not enabled or the authentication fails, the port still uses the guest VLAN to access limited network resources.
- If the authentication server fails to authenticate the client (for example, an incorrect user name or password is entered), all services on this port go through the Restrict VLAN. The Restrict VLAN has similar functions as the guest VLAN. Both allow users to access limited network resources before passing authentication. However, the Restrict VLAN has fewer network resources than the guest VLAN. Therefore, the Restrict VLAN more strictly limits the network resources accessible to unauthenticated users.
- If the authentication server cannot be connected (for example, the network between the device and the authentication server is disconnected or the authentication server is faulty), all services on the port can go through the Critical VLAN, for example, emergency voice call or service assurance.

Figure 17-2 shows the system authentication process.

Figure 17-2 Authentication process



 Copyright © Huawei Technologies Co., Ltd.

NOTE

The relationship between MAB authentication and port-based 802.1X authentication is as follows: If both port-based 802.1X authentication and MAB authentication are enabled, 802.1X authentication is performed first. If 802.1x authentication times out, MAB authentication is then performed. If 802.1x authentication is disabled but MAB authentication is enabled, MAB authentication is directly performed. Disabling MAB authentication has no impact on 802.1X authentication status.

17.4 802.1X Information Exchange

Authorized/Unauthorized 802.1X Port

Whether the client is authorized to access the network depends on the status of the connected device port. The port is started in the unauthorized state. In this state, all incoming and outgoing traffic except 802.1X protocol packets is not allowed to pass through the port. After the client passes the authentication, the port changes to the authorized state to allow all the traffic from the client.

If a client that does not support 802.1X authentication is connected to an unauthorized 802.1X port, the device requests to authenticate the client. In this case, the client does not respond to the request, the port is still in the unauthorized state, and the client does not have the network access right. The **port dot1x port-control** command and the following keywords can be used to control the port authorization status of the device:

force-authorized: Specifies the forced authorization mode. In this mode, the port always stays in the authorized state and users can access the network through this port without authentication or authorization. The port sends and receives traffic normally without 802.1X client authentication. This configuration has security risks and is not recommended.

force-unauthorized: Specifies the forced unauthorization mode. The port stays in the unauthorized state and users are not allowed to access network resources through this port.

auto: Specifies the automatic identification mode. After 802.1X authentication is enabled, the port is in the unauthorized state and can send and receive only EAPOL packets. If the port receives an EAPOL-start frame, authentication starts. The device requests the identity of the client and starts to relay the authentication message between the client and the authentication server. If the authentication is successful (the client receives an Accept frame from the authentication server), the port status changes to authorized and all frames from the authenticated client are allowed to pass through the port. If the authentication fails, the port remains in the unauthorized state.

802.1X Information Exchange

EAP packets are exchanged between the device and authentication server in EAP relay mode or EAP termination mode. Therefore, 802.1X port-based authentication can be performed in either of the following modes: EAP relay and EAP termination.

EAP Relay

 Copyright © Huawei Technologies Co., Ltd.

802.1x authentication can be triggered by the client or the device. If the **port dot1x port-control auto** command is used to enable identity authentication on the port, the device sends an EAP-Request/Identity frame to the client when detecting that the link is connected (the link status changes from down to up). After receiving the frame, the client responds with an EAP-response/identity frame.

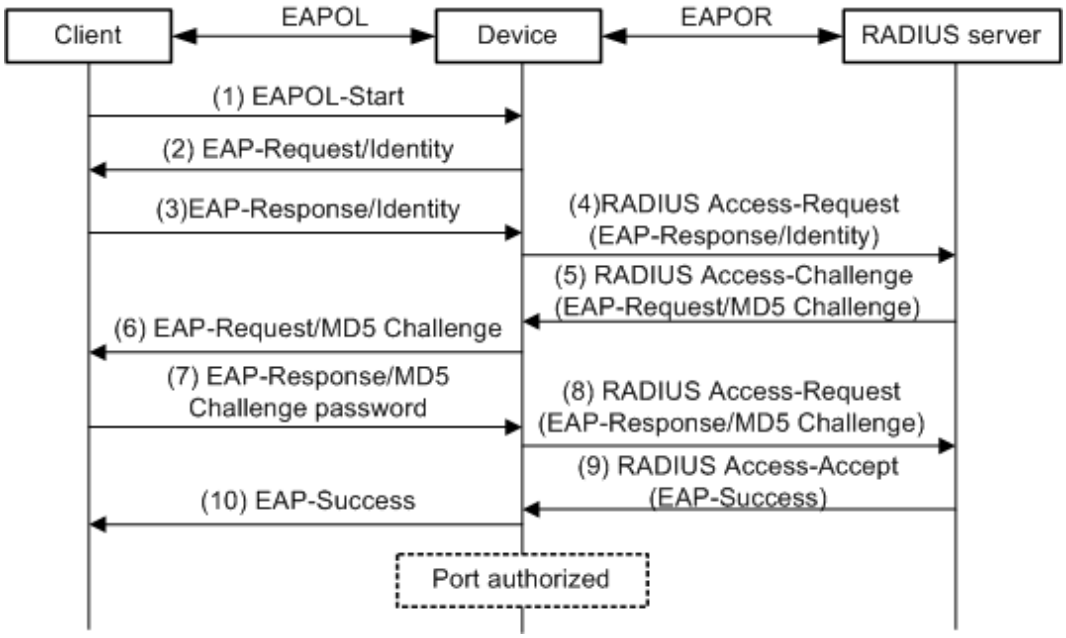
If the client does not receive the EAP-Request/Identity frame from the device during startup, the client can send an EAPOL-start frame to start authentication. This frame can instruct the device to send a request.

When the client provides its identity, the device acts as a mediator to transfer the EAP frame between the client and the authentication server until the authentication succeeds or fails. If the authentication succeeds, the port is authorized. If the authentication fails, it can be performed again. Based on the final authentication result, the port is assigned to a VLAN that provides limited services or the user is not assigned the network access right.

The device relays EAP packets between the client and RADIUS server:

- Re-encapsulates EAPOR packets into EAPOL packets and sends them to the client.
- Re-encapsulates EAPOL packets into EAROR packets and sends them to the RADIUS server.

Figure 17-3 EAP relay authentication process



(1) When a user accesses the network, the 802.1X client program is automatically started. The client program sends an authentication request frame (EAPOL-start) to the device to start the authentication process.

(2) After receiving the authentication request frame from the client, the device sends an EAP-Request/Identity frame to the client, requesting the client to send the user name.

(3) The client sends the user name to the device through an EAP-Response/Identity frame to respond to the request sent by the device.

 Copyright © Huawei Technologies Co., Ltd.

(4) The device encapsulates the EAP packets from the client in EAPOR format into a RADIUS Access-Request packet and sends it to the authentication server for processing.

(5)–(9) The authentication server (RADIUS server) verifies the user name and password of the client using a specific method. If the authentication succeeds, the server sends an authentication success packet (RADIUS Access-Accept) to the device.

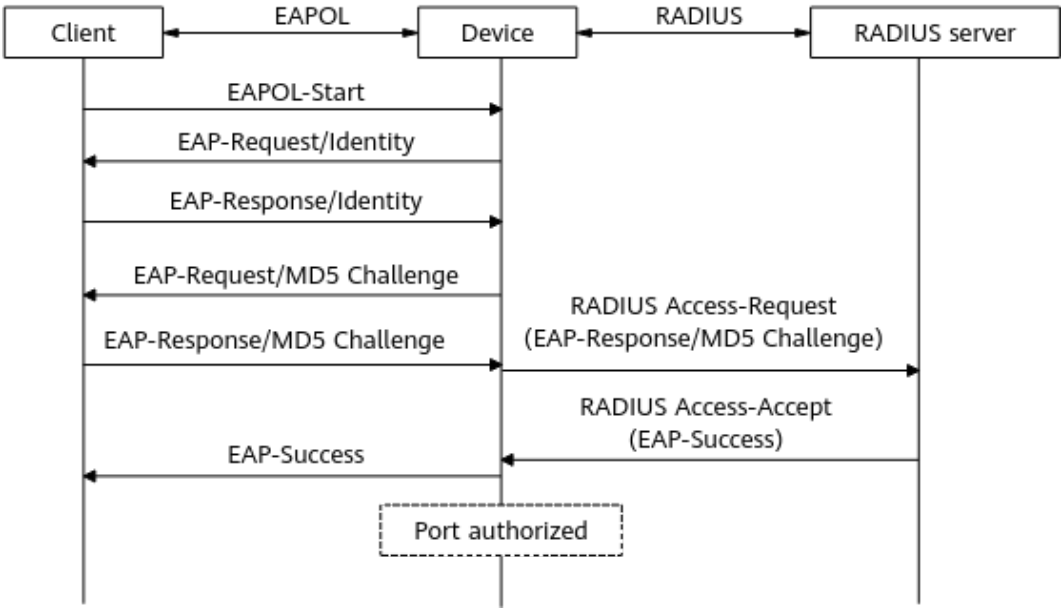
(10) After receiving the RADIUS Access-Accept packet, the device decapsulates it in EAPOR format and sends an EAP-Success packet to the client. Then, the device changes the port to the authorized state and allows the user to access the network through the port.

EAP Termination

The device terminates the EAP authentication process:

- Encapsulates the client's authentication information in received EAP packets into standard RADIUS packets.
- Communicates with the RADIUS server using the Password Authentication Protocol (PAP) or Challenge Handshake Authentication Protocol (CHAP) for access authentication.

Figure 17-4 EAP termination authentication process



The difference between the 802.1X authentication process in EAP termination mode and that in EAP relay mode is as follows: In EAP termination mode, the device randomly generates an MD5-Challenge value for user password encryption, and then sends the user name, random MD5-Challenge value, and password encrypted on the client to the RADIUS server for authentication.

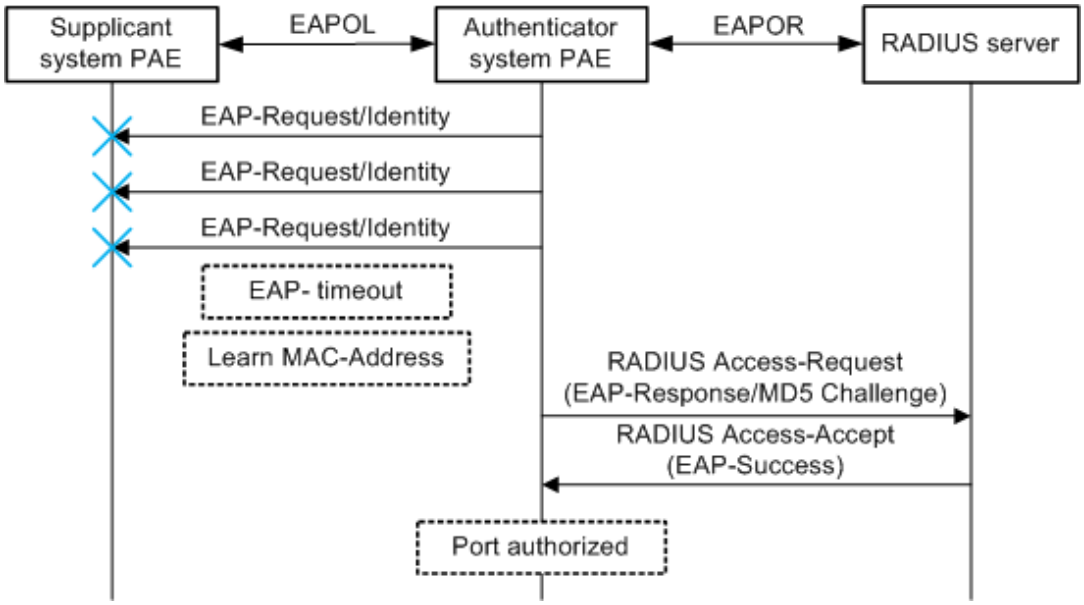
MAC Bypass (MAB) Authentication

If MAB authentication is enabled and 802.1X authentication times out during EAPOL message exchange, the device can authorize the client when detecting

 Copyright © Huawei Technologies Co., Ltd.

Ethernet packets from the client. The device uses the MAC address of the client as its identifier, adds the MAC address to the RADIUS-Access/Request frame, and sends the frame to the RADIUS server. After the server sends a RADIUS-Access/Accept frame (authorization success) to the device, the port is authorized. If the MAB authorization fails and the guest VLAN is specified, the device assigns the port to the guest VLAN. If the device detects an EAPOL packet when waiting for an Ethernet packet, the device stops MAB authorization and starts 802.1X authentication. The detailed process is shown in **Figure 17-5**. In the figure, PAE refers to the port access entity, that is, the device for authentication and authorization.

Figure 17-5 MAB authentication information exchange process



Re-authentication and Forcing Users to go Offline

If an administrator modifies the access rights and authorization parameters (such as the dynamic service VALN) of an online user, re-authentication is required for the user.

After the re-authentication function is configured for online 802.1X users, the device sends the saved authentication parameters of an online user to the authentication server for re-authentication (when a user goes online, the device saves the user authentication information). If the authentication information of the user on the authentication server remains unchanged, the user keeps online. If the authentication information has been changed, the user is logged out and needs to be re-authenticated.

RADIUS CoA

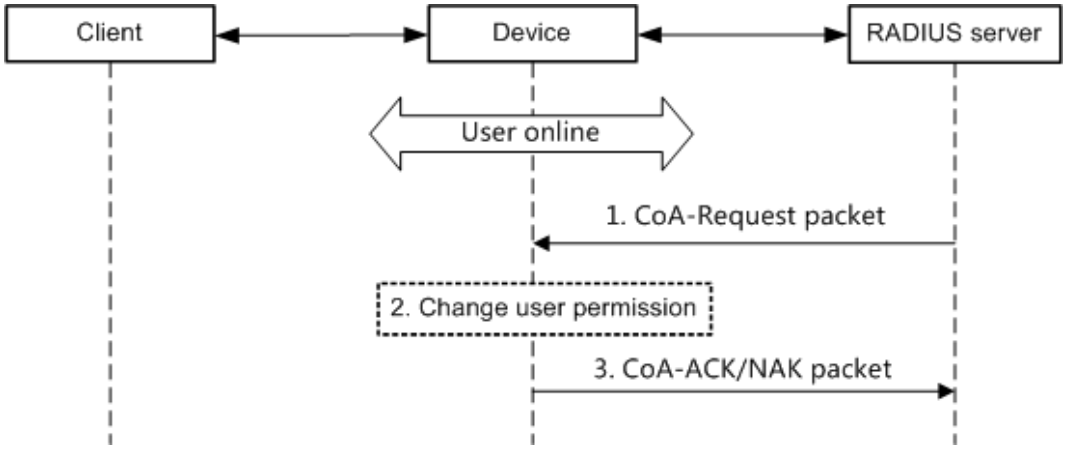
Change of Authorization (CoA) allows the administrator to modify the rights of online users or re-authenticate the users through RADIUS after they are successfully authenticated. **Figure 17-6** shows the CoA packet exchange process.

 Copyright © Huawei Technologies Co., Ltd.

Table 17-1 RADIUS CoA packets

Packet Name	Description
CoA-Request	Indicates a dynamic authorization request packet. When the administrator needs to modify the rights of an online user (for example, the administrator does not want the user to access a website), the RADIUS server can send a dynamic authorization request packet to the RADIUS client to enable the client to modify the rights of the online user.
CoA-ACK	Indicates a dynamic authorization accept packet. If the RADIUS client successfully modifies the user rights, the RADIUS client sends a dynamic authorization accept packet to the RADIUS server.
CoA-NAK	Indicates a dynamic authorization reject packet. If the RADIUS client fails to modify the user rights, the RADIUS client sends a dynamic authorization reject packet to the RADIUS server.

Figure 17-6 CoA packet exchange process



1. The RADIUS server sends a CoA-Request packet to the device based on the service information, requesting the device to modify the user authorization information. The packet may include authorization information such as an ACL rule.
2. After receiving the CoA-Request packet, the device matches the packet with user information on the device to identify the user. If the matching is successful, the authorization information of the user is modified. If the matching fails, the original authorization information of the user is retained.

The device replies with a CoA-ACK or CoA-NAK packet.

- If the change is successful, the device returns a CoA-ACK packet to the RADIUS server.
- If the change fails, the device returns a CoA-NAK packet to the RADIUS server.

 Copyright © Huawei Technologies Co., Ltd.

The following table lists the Huawei extended RADIUS attributes involved in this feature.

Table 17-2 Huawei extended RADIUS attributes involved in RADIUS CoA

Attribute ID	Attribute Name	Attribute Type	Description
26-238	HW-Ext-Specific	string	User extension attribute. • user-command: The value can be 1, 2, or 3. – 1: Re-authenticates a user. In this case, you need to set this attribute as user-command=1 on the authentication server. **NOTE** If user-command is set to 1, other authorization attributes are not supported. – 2: The authentication port is intermittently disconnected. In this case, you need to set this attribute as user-command=2 on the authentication server. – 3: The authentication port is disabled. In this case, you need to set this attribute as user-command=3 on the authentication server. **NOTE** • You can run the **display ont port state** *portid ontid* **eth-port all** command to query the status of an ONU port. • After the authentication port is disabled, you can run the **ont port attribute** *portid ontid* **eth** *ont-portid* **operational-state** *switch* command to restore the port to the normal state.

RADIUS DM

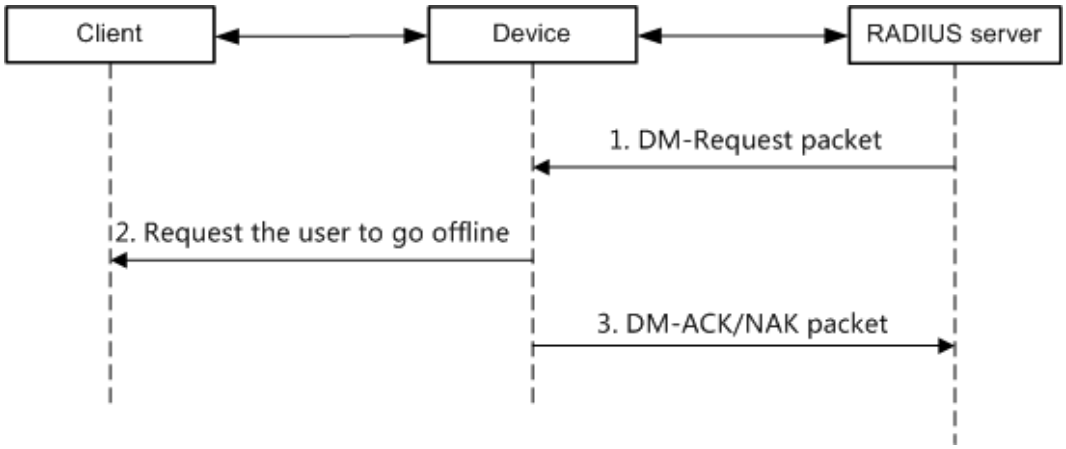
A Disconnect Message (DM) is sent by a RADIUS server to force a user to go offline. **Figure 17-7** shows the DM packet exchange process.

 Copyright © Huawei Technologies Co., Ltd.

Table 17-3 RADIUS DM packets

Packet Name	Description
DM-Request	Indicates a user logout request packet. To log out an online user, the administrator can configure the RADIUS server to send a user offline request packet to the RADIUS client. The RADIUS client then terminates the user connection.
DM-ACK	Indicates a user logout request accept packet. After terminating the user connection, the RADIUS client sends a user logout request accept packet to the RADIUS server.
DM-NAK	Indicates a user logout request reject packet. If the RADIUS client fails to disconnect the user, the RADIUS client sends a user logout request reject packet to the RADIUS server.

Figure 17-7 DM packet exchange process



1. After the administrator forces a user to go offline on the RADIUS server, the RADIUS server sends a DM-Request packet to the device to request the user to go offline.
2. After receiving the DM-Request packet, the device matches the packet with user information on the device to identify the user. If the matching is successful, the device instructs the user to go offline. If the matching fails, the user remains online.
3. The device replies with a DM-ACK or DM-NAK packet.
   - If the user goes offline successfully, the device sends a DM-ACK packet to the RADIUS server.
   - If the user does not go offline, the device sends a DM-NAK packet to the RADIUS server.

Compared with user online authorization or active user offline processes, the CoA/DM request packet is sent by the server, and the CoA/DM response packet is sent by the device in CoA/DM scenarios. If the operation is successful, the device replies with an ACK packet. If the operation fails, the device replies with a NAK packet.

 Copyright © Huawei Technologies Co., Ltd.

Active RADIUS Server Detection

The reachability and availability of the RADIUS server are the basic conditions for user access authentication. If the authentication server cannot be connected (for example, the network between the device and the authentication server is disconnected or the authentication server is faulty), the RADIUS server cannot authenticate or authorize users. To solve this problem, the device supports the escape function when the RADIUS server is down. That is, all services on a port can access resources in a critical VLAN, for example, in emergency voice call or service assurance scenarios. However, the escape function can be enabled only after the device sets the RADIUS server status to Down. If the status of the RADIUS server is not set to Down and the device cannot communicate with the RADIUS server, users cannot obtain authorization from the server or escape, and cannot access network resources for a long time. Therefore, the device must detect the RADIUS server status in a timely manner so that users can immediately obtain the escape right when the RADIUS server is down. After the RADIUS server goes up, users exit the escape mode and are re-authenticated.

Server Down/Up State Transition

By default, the server is in the Up state. It is marked Down when the detection function is enabled and a new user who initiates an authentication request does not receive any response from the server.

If either the primary or secondary server is down, the device starts to detect the server status at the configured detection interval by sending RADIUS probe packets to the server. After the server returns a response packet, the server status changes to Up and the detection stops.

If both the primary and secondary servers are down, the device does not forward authentication request packets initiated by a user to the server. Instead, the device directly terminates the authentication process and allocates a critical VLAN to the user to enable the escape function. After either of the primary or secondary server goes up, the device sends user authentication request packets to the server to start the authentication process.

ACL

Access control list (ACL) is a widely used tool for filtering network traffic. An ACL can identify specific traffic on the network based on the source IP address, destination IP address, protocol type, and source and destination ports in packets, and then process the traffic accordingly.

On the ONU, different ACLs can be configured to filter different types of users, and then implement different access control policies for the users.

Pre-auth ACL

When a user client connects to the network but does not initiate 802.1X authentication, a pre-auth ACL can be configured to filter user packets that meet certain conditions (such as the specified destination IP address and destination MAC address) so that the user can connect to the specified network, for example, logging in to a website. To use a pre-auth ACL, the system administrator needs to pre-configure ACL rules on the ONU.

The pre-auth ACL takes effect only in the period from the time when the user fails to be authenticated, the time when the RADIUS server is unreachable, and the

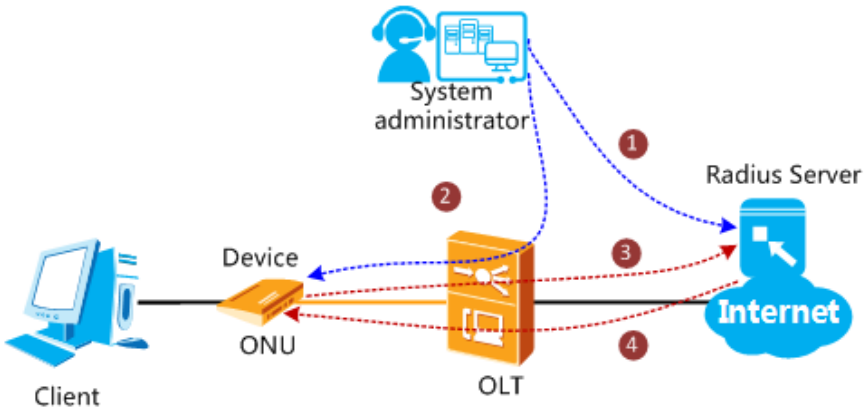
 Copyright © Huawei Technologies Co., Ltd.

time when the user initiates authentication to the time when the authentication succeeds.

Dynamic ACL

After a user passes 802.1X authentication, the RADIUS server sends a RADIUS packet carrying dynamic ACL information to the network access server (NAS) and applies the ACL rule to the user, as shown in **Figure 17-8**.

Figure 17-8 Dynamic ACL process



1. The system administrator configures the name of the dynamic ACL rule that takes effect after user authentication on the RADIUS server.
2. The system administrator configures a dynamic ACL rule on the ONU through the OLT.
3. The ONU initiates 802.1X authentication to the RADIUS server.
4. The RADIUS server sends the authentication result and dynamic ACL rule name to the ONU, and the ONU enables the specified dynamic ACL rule to take effect.

The dynamic ACL name is carried in the RADIUS standard attribute field Filter-ID, as shown in **Table 17-4**. In addition, specific rules need to be configured on the device.

Table 17-4 Standard RADIUS attributes involved in dynamic ACL

Attribute ID	Attribute Name	Attribute Type	Description
11	Filter-ID	string	Indicates the name of an IPv4 ACL for a user.

Download ACL

Download ACL is another type of ACL on the RADIUS server. You can configure download ACL rules in plaintext. It also applies to users who are successfully authenticated.

The configuration information of a download ACL is carried by Huawei extended RADIUS attributes, as listed in **Table 17-5**.

 Copyright © Huawei Technologies Co., Ltd.

Table 17-5 Huawei extended RADIUS attributes involved in the download ACL feature

Attribute ID	Attribute Name	Attribute Type	Description
26-82	HW-Data-Filter	string	The RADIUS server uses this attribute to deliver ACL rules to users. **NOTE** • A RADIUS packet can carry multiple No. 26–82 attributes. Currently, each attribute can carry only one ACL rule. • You can run the **display access-user user-id** *user-id* command to check whether this attribute takes effect. If the command output displays **Dynamic ACL desc (Effective)**, this attribute takes effect. If **Dynamic ACL desc (Ineffective)** is displayed, the attribute does not take effect. **New attribute format (fields in square brackets are optional)** Format: $*number* **permit**/**deny** [*protocol*] [*direction ip-address* [*port*]] The fields are described as follows: • $: indicates the start character of each ACL rule. • *number*: specifies the last 3 digits of an ACL rule number. The value ranges from 0 to 999. The first 2 digits of an ACL rule number are fixed to 10. For example, if an ACL rule number is 12, the actual ACL rule number is 10012. • **permit**/**deny**: indicates the ACL action. **permit** indicates that access is permitted, and **deny** indicates that access is denied. • *protocol*: specifies the protocol type. The value can be tcp or udp. • *direction*: indicates the IP address type. The value can be dst or src. dst indicates the destination IP address, and src indicates the source IP address. • *ip-address*: specifies an IP address. The value can be any or x.x.x.x/xx. x.x.x.x indicates an IP address in dotted decimal notation and xx indicates the mask in decimal notation. The IP address and mask are separated by a slash (/). • *port*: indicates a port number. Currently, only one port is supported. Example:

 Copyright © Huawei Technologies Co., Ltd.

Attribute ID	Attribute Name	Attribute Type	Description
			• **$1 permit dst 10.0.239.192/26** • **$2 permit udp src any 8080** • **$5 deny** **Old Attribute Format** **acl** *number* **key1** *key-value1*... **keyN** *key-valueN* **permit**/**deny** The fields are described as follows: • **acl**: indicates that ACL rules are delivered. • *number*: specifies the number of an ACL rule. The value ranges from 10000 to 10999. • **keyM** *key-valueM*(1≤M≤N): indicates the keyword and value of an ACL rule. The keyword can be: – **dest-ip** *ip-address*: specifies the destination IP address. The value is in dotted decimal notation. When the destination IP address is 0.0.0.0, this parameter does not need to be configured. – **dest-ipmask** *mask*: specifies the destination IP address mask. For NAC users, the value is an integer ranging from 0 to 32. For VM users, the value is in dotted decimal notation. If the IP address mask is 0, you do not need to set this parameter. – **tcp-srcport** *port*: specifies the source TCP start port number. The value ranges from 0 to 65535. – **tcp-srcport-end** *port*: specifies the source TCP end port number. The value ranges from 0 to 65535 and must be greater than the start port number. – **tcp-dstport** *port*: specifies the destination TCP start port number. The value ranges from 0 to 65535. – **tcp-dstport-end** *port*: specifies the destination TCP end port number. The value ranges from 0 to 65535 and must be greater than the start port number. – **udp-srcport** *port*: specifies the source UDP start port number. The value ranges from 0 to 65535.

 Copyright © Huawei Technologies Co., Ltd.

Attribute ID	Attribute Name	Attribute Type	Description
			– **udp-srcport-end** *port*: specifies the source UDP end port number. The value ranges from 0 to 65535 and must be greater than the start port number. – **udp-dstport** *port*: specifies the destination UDP start port number. The value ranges from 0 to 65535. – **udp-dstport-end** *port*: specifies the destination UDP end port number. The value ranges from 0 to 65535 and must be greater than the start port number. ● **permit**/**deny**: indicates the ACL action. **permit** indicates that access is permitted, and **deny** indicates that access is denied. **NOTE** ● It is recommended that the standard RADIUS attribute Filter-Id be used to deliver ACLs. ● A maximum of 64 ACL rules can be delivered to a single user. If too many ACL rules are delivered, the number of online users and the access rate will be affected. It is recommended that no more than 16 ACL rules be delivered to a single user. ● When the direct forwarding mode is used, ACL rules cannot be delivered through this attribute to wireless users. Instead, ACL rules can be delivered through standard RADIUS attribute 11. ● All keywords are case insensitive. All keywords and (or) key values are separated by spaces. The sequence of keywords can be flexible. The permit and deny keywords can be placed after *number* or at the end of the statement. Example: ● **acl 10005 deny** ● **acl 10006 tcp-dstport 5080 permit** ● **acl 10007 dest-ip 10.11.11.2 dest-ipmask 32 permit** ● **acl 10008 dest-ip 10.11.11.3 dest-ipmask 32 udp-dstport 5070 permit** ● **acl 10009 dest-ip 11.1**

Automatic Logout After Idle Time Expires

When the idle time of a user exceeds the specified time, the user is automatically logged out. You can configure the standard RADIUS attribute field Idle-Timeout on the server to implement this function.

 Copyright © Huawei Technologies Co., Ltd.

Table 17-6 Standard RADIUS attributes

Attribute ID	Attribute Name	Attribute Type	Description
28	Idle-Timeout	integer	Indicates the maximum idle time of a user before a session ends, that is, the idle logout time of the user, in seconds.

17.5 Configuring 802.1X Authentication

This section describes how to configure 802.1X authentication.

17.5.1 ONU Functioning as the 802.1x Authentication Device

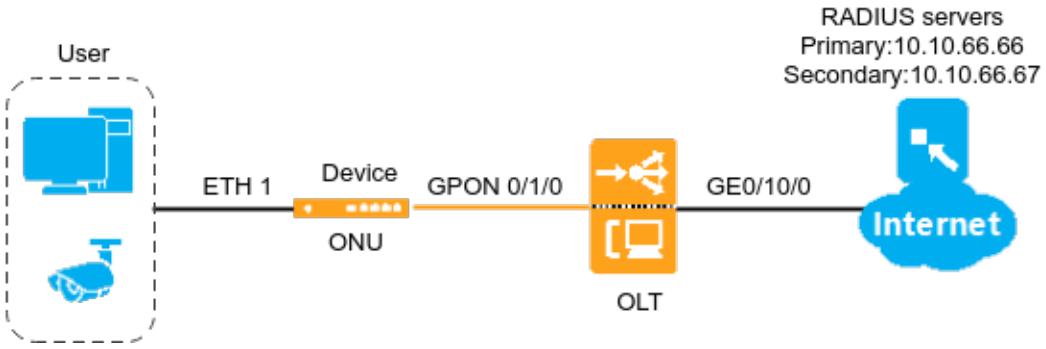
The ONU functions as the 802.1X authentication device. 802.1x authentication configurations are delivered to the ONU through the OLT.

Prerequisites

- The OLT has been connected to a Layer 3 switch.
- The ONU is online and can be managed by the OLT.
- The remote RADIUS server has been deployed, and the network connection is normal.

Figure 17-9 shows the networking diagram for RADIUS authentication.

Figure 17-9 Networking diagram for RADIUS authentication



Data Planning

Configuration Item	Value
Networking data	PON port ID: 0/1/0 Upstream port: 0/10/0 ONU ID: 0 ONU IP: 10.10.10.20

 Copyright © Huawei Technologies Co., Ltd.

Configuration Item	Value
802.1X authentication	Port authentication mode: auto Authentication method: EAP ONU port authentication granularity: MAC address authentication **NOTE** Port-based authentication and MAC address-based authentication are applicable to the following scenarios: • In MAC-based authentication mode, every 802.1X user accessing an interface must be authenticated. If a user goes offline, other authenticated users can still access the network. This authentication mode applies to scattered users. • In port-based authentication mode, other 802.1X users can use network resources without being authenticated as long as one user on the interface is successfully authenticated. However, if the authenticated user goes offline, other users cannot use network resources. This authentication mode applies to group users. • By default, the port-based authentication mode is used.
VLAN planning	• VLAN ID of the management WAN port: 11 • Guest VLAN, SVLAN: 10; CVLAN: 10 • Restrict VLAN, SVLAN: 20; CVLAN: 20 • Critical VLAN, SVLAN: 30; CVLAN: 30 • Dynamic Service VLAN, SVLAN: 40; CVLAN: 40 **NOTE** The preceding VLANs must be mapped to GEM ports in the line profile and service profile during network planning.
RADIUS server	• Primary authentication server: IP address: 10.10.66.66; Port number: 1812 • Secondary authentication server: IP address: 10.10.66.67; Port number: 1812 • Shared primary key: 0123456789123456 • Shared secondary key: 0123456789123456 • The IP address and shared key of the charging and authorization servers are the same as those of the authentication server.

Procedure

Step 1 Create and configure the dot1x (802.1X) profile.

Set the ONU port authentication mode to MAC-based authentication, enable dot1x authentication, set the port authentication mode to auto, set the

 Copyright © Huawei Technologies Co., Ltd.

authentication method to EAP, and configure the related VLAN. If ONU port 1 connects to a terminal that does not support 802.1X, MAB authentication needs to be enabled.

NOTE

The ONU port authentication granularity can be set to MAC address authentication only when the ONU supports the MAC address authentication granularity. Otherwise, the ONU configuration fails.

```
huawei(config)#ont dot1x-profile profile-id 1
huawei(config-dot1x-profile-1)#port dot1x eth 1 enable
huawei(config-dot1x-profile-1)#port dot1x authentication-method eth 1 eap
huawei(config-dot1x-profile-1)#port dot1x port-control eth 1 auto
huawei(config-dot1x-profile-1)#port dot1x port-method eth 1 mac
huawei(config-dot1x-profile-1)#port dot1x guest-vlan eth 1 10
huawei(config-dot1x-profile-1)#port dot1x restrict-vlan eth 1 20
huawei(config-dot1x-profile-1)#port dot1x critical-vlan eth 1 30
huawei(config-dot1x-profile-1)#port mac-bypass eth 1 enable
huawei(config-dot1x-profile-1)#quit
```

Step 2 Configure a WAN profile.

```
huawei(config)#ont wan-profile profile-id 9 profile-name wan_prof_hwtest
huawei(config-wan-profile-9)#connection-type route
huawei(config-wan-profile-9)#quit
```

Step 3 Configure the IP address of the ONU and the IP interface index of the Internet service, and bind the WAN profile.

Set the static IP address of the ONU to 10.10.10.20 and the IP interface index of the Internet service to 0.

```
huawei(config)#gpon ont home-gateway config-method omci
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#ont ipconfig 0 0 ip-index 0 static ip-address 10.10.10.20 mask
255.255.255.0 vlan 11 gateway 10.10.10.10
huawei(config-if-gpon-0/1)#ont internet-config 0 0 ip-index 0
huawei(config-if-gpon-0/1)#ont wan-config 0 0 ip-index 0 profile-id 9
huawei(config-if-gpon-0/1)#quit
```

Step 4 Configure the RADIUS server IP address and authentication domain.

Configure the RADIUS server 10.10.66.66 as the primary authentication server, the RADIUS server 10.10.66.67 as the secondary authentication server, and configure the primary and secondary shared keys as 0123456789123456. Set the authentication timeout interval to 20s and the maximum number of retries to 3. RADIUS authentication servers work in primary/secondary mode.

```
huawei(config-dot1x-profile-1)#radius-server authentication 10.10.66.66 1812
huawei(config-dot1x-profile-1)#radius-server shared-key 0123456789123456
huawei(config-dot1x-profile-1)#radius-server authentication 10.10.66.67 1812 secondary
huawei(config-dot1x-profile-1)#radius-server shared-key 0123456789123456 secondary
huawei(config-dot1x-profile-1)#radius-server user-name domain-included
huawei(config-dot1x-profile-1)#radius-server accounting 10.10.66.66 1813
huawei(config-dot1x-profile-1)#radius-server accounting 10.10.66.67 1813 secondary
huawei(config-dot1x-profile-1)#radius-server accounting shared-key 0123456789123456
huawei(config-dot1x-profile-1)#radius-server accounting shared-key 0123456789123456 secondary
huawei(config-dot1x-profile-1)#radius-server authorization 10.10.66.66 shared-key 0123456789123456
huawei(config-dot1x-profile-1)#radius-server authorization 10.10.66.67 shared-key 0123456789123456
huawei(config-dot1x-profile-1)#radius-server timeout 20
huawei(config-dot1x-profile-1)#radius-server retransmit 3
huawei(config-dot1x-profile-1)#radius-server authentication mode backup
huawei(config-dot1x-profile-1)#quit
```

Step 5 Configure service flows for 802.1X authentication.

 Copyright © Huawei Technologies Co., Ltd.

- Add a service flow to the RADIUS server. Set SVLAN and CVLAN to 11.
  ```
  huawei(config)#vlan 11 smart
  huawei(config)#port vlan 11 0/10 0
  huawei(config)#service-port vlan 11 gpon 0/1/0 ont 0 gemport 0 multi-service user-vlan 11
  ```
- Add a service flow to the Guest network. Set SVLAN and CVLAN to 10.
  ```
  huawei(config)#vlan 10 smart
  huawei(config)#port vlan 10 0/10 0
  huawei(config)#service-port vlan 10 gpon 0/1/0 ont 0 gemport 0 multi-service user-vlan 10
  ```
- Add a service flow to the Restrict network. Set SVLAN and CVLAN to 20.
  ```
  huawei(config)#vlan 20 smart
  huawei(config)#port vlan 20 0/10 0
  huawei(config)#service-port vlan 20 gpon 0/1/0 ont 0 gemport 0 multi-service user-vlan 20
  ```
- Add a service flow to the Critical network. Set SVLAN and CVLAN to 30.
  ```
  huawei(config)#vlan 30 smart
  huawei(config)#port vlan 30 0/10 0
  huawei(config)#service-port vlan 30 gpon 0/1/0 ont 0 gemport 0 multi-service user-vlan 30
  ```
- Add a service flow to the Dynamic Service network. Set SVLAN and CVLAN to 40.
  ```
  huawei(config)#vlan 40 smart
  huawei(config)#port vlan 40 0/10 0
  huawei(config)#service-port vlan 40 gpon 0/1/0 ont 0 gemport 0 multi-service user-vlan 40
  huawei(config)#save
  ```

Step 6 (Optional) Configure the Pre-auth ACL function of the ONU port.

- Configure the pre-auth ACL for ONU port 1 to allow the user packets with the destination IP address 10.10.100.1 to pass through.
  ```
  huawei(config)#ont acl-profile profile-id 1
  huawei(config-acl-profile-1)#acl add acl-id 1 acl-name acl-preauth
  huawei(config-acl-profile-1)#rule add acl-id 1 rule-id 0 priority 1 permit
  huawei(config-acl-profile-1)#rule add acl-id 1 rule-id 1 priority 1 deny
  huawei(config-acl-profile-1)#conditon add acl-id 1 rule-id 0 condition-id 0 validation-operator equal field dst-ipv4-address 10.10.100.1
  huawei(config-acl-profile-1)commit
  huawei(config)#interface gpon 0/1
  huawei(config-if-gpon-0/1)#ont acl-config 0 0 profile-id 1
  huawei(config-if-gpon-0/1)#quit
  huawei(config)#ont dot1x-profile profile-id 1
  huawei(config-dot1x-profile-1)#port dot1x acl-id eth 1 1
  huawei(config-dot1x-profile-1)#quit
  ```

Step 7 (Optional) Configure the dynamic ACL function for ONU users.

- Configure a dynamic ACL for the ONU to allow user packets with destination MAC address 00e0-fc00-0033 to pass through.
  ```
  huawei(config)#ont acl-profile profile-id 1
  huawei(config-acl-profile-1)#acl add acl-id 2 acl-name acl-dynamic.in
  huawei(config-acl-profile-1)#rule add acl-id 2 rule-id 0 priority 1 permit
  huawei(config-acl-profile-1)#conditon add acl-id 2 rule-id 0 condition-id 0 validation-operator equal field dst-mac-address 00e0-fc00-0033
  huawei(config-acl-profile-1)commit
  huawei(config)#interface gpon 0/1
  huawei(config-if-gpon-0/1)#ont acl-config 0 0 profile-id 1
  huawei(config-if-gpon-0/1)#quit
  ```

Step 8 (Optional) Configure the ONU to automatically detect the RADIUS server.

Use the specified user name and password to detect the RADIUS authentication server at an interval of 30 seconds.

```
huawei(config)#ont dot1x-profile profile-id 1
huawei(config-dot1x-profile-1)radius-server auto-detect username user@2020 password
huawei(config-dot1x-profile-1)radius-server auto-detect interval 30
huawei(config-dot1x-profile-1)quit
```

Step 9 Bind the ONU to the dot1x profile.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#ont dot1x-config 0 0 profile-id 1
huawei(config-if-gpon-0/1)#quit
```

----End

Result

After the ONU is configured, it can be used as an access point for 802.1X authentication. A connected terminal can access the LAN by using the user name and password, or by using the MAC address.

17.5.2 OLT Functioning as the 802.1X Authentication Device (Configured on the CLI).

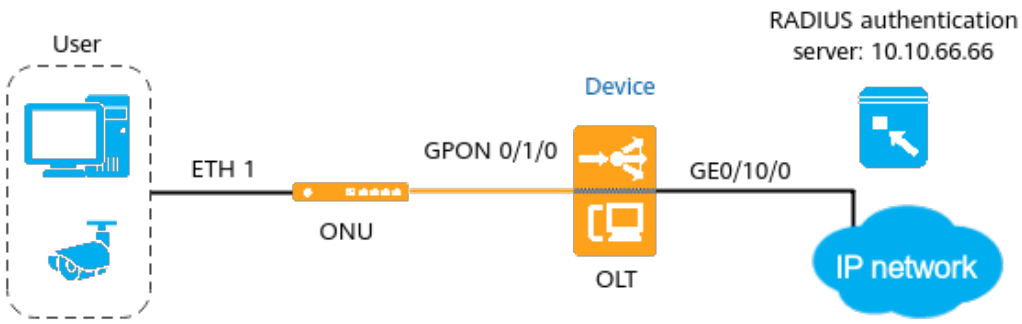
As the 802.1X authentication device, an OLT authenticates clients and controls the network access rights or scope of the clients.

Prerequisites

The remote RADIUS server has been deployed, and the network connection is normal.

Figure 17-10 shows the networking diagram for RADIUS authentication.

Figure 17-10 Networking diagram for RADIUS authentication



Data Planning

Configuration Item	Value
Networking data	PON port: 0/1/0 Upstream port: 0/10/0 ONU ID: 0
802.1X authentication	Service flow ID: 1 Service VLAN ID for MAC address–based 802.1X authentication: 20 Enabling MAC bypass authentication
VLAN planning	User side VLAN: 100

 Copyright © Huawei Technologies Co., Ltd.

Configuration Item	Value
RADIUS server	• Authentication server IP address: 10.10.66.66; port number: 1812 • Shared key: 0123456789123456

Procedure

Step 1 Configure a GPON ONU profile. The line profile is configured with the 802.1p priority mapping and the service profile is configured with the transparent transmission mode. In this way, packets of all VLANs on ONUs can pass through, and service flows can be dynamically created when the RADIUS server issues any dynamic VLAN.

1. Configure a DBA profile.

   Run the **display dba-profile** command to query existing DBA profiles in the system. If existing DBA profiles in the system do not meet the requirement, run the **dba-profile add** command to create a DBA profile.

   Add a DBA profile with service profile ID 10 and type 4.

   ```
   huawei(config)#dba-profile add profile-id 10 type4 max 1024000
   ```

2. Configure an ONU line profile.

   Add GPON ONU line profile 10 and bind T-CONT 1 to DBA profile 10. In this way, the T-CONT provides flexible DBA solutions based on different configurations in the DBA profile.

   ```
   huawei(config)#ont-lineprofile gpon profile-id 10
   huawei(config-gpon-lineprofile-10)#tcont 1 dba-profile-id 10
   ```

   Use GEM port 0 to carry services and bind the GEM port to a T-CONT.

   ```
   huawei(config-gpon-lineprofile-10)#gem add 0 eth tcont 1
   ```

   Set the mapping mode between GEM ports and the ONU-side services to 802.1p priority mapping.

   ```
   huawei(config-gpon-lineprofile-10)#mapping-mode priority
   huawei(config-gpon-lineprofile-10)#gem mapping 0 0 priority 0
   huawei(config-gpon-lineprofile-10)#gem mapping 0 1 priority 1
   huawei(config-gpon-lineprofile-10)#gem mapping 0 2 priority 2
   huawei(config-gpon-lineprofile-10)#gem mapping 0 3 priority 3
   huawei(config-gpon-lineprofile-10)#gem mapping 0 4 priority 4
   huawei(config-gpon-lineprofile-10)#gem mapping 0 5 priority 5
   huawei(config-gpon-lineprofile-10)#gem mapping 0 6 priority 6
   huawei(config-gpon-lineprofile-10)#gem mapping 0 7 priority 7
   ```

   After the configuration is completed, run the **commit** command to make the configured parameters take effect.

   ```
   huawei(config-gpon-lineprofile-10)#commit
   huawei(config-gpon-lineprofile-10)#quit
   ```

3. Configure an ONU service profile.

   Set the profile ID to 100. Set the maximum number of auto-adaptive ETH ports on an ONU to 8 in the ONU service profile. The port capability set profile of an ONU must be the same as the actual capability set of the ONU. Set the packet forwarding mode of a port to TAG transparent transmission. In this mode, the port does not process the packets and keeps the VLAN tag unchanged.

   ```
   huawei(config)#ont-srvprofile gpon profile-id 10
   huawei(config-gpon-srvprofile-10)#ont-port eth adaptive 8
   ```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-gpon-srvprofile-10)#port vlan eth 1 transparent
huawei(config-gpon-srvprofile-10)#port vlan eth 2 transparent
huawei(config-gpon-srvprofile-10)#port vlan eth 3 transparent
huawei(config-gpon-srvprofile-10)#port vlan eth 4 transparent
huawei(config-gpon-srvprofile-10)#port vlan eth 5 transparent
huawei(config-gpon-srvprofile-10)#port vlan eth 6 transparent
huawei(config-gpon-srvprofile-10)#port vlan eth 7 transparent
huawei(config-gpon-srvprofile-10)#port vlan eth 8 transparent
huawei(config-gpon-srvprofile-10)#commit
huawei(config-gpon-srvprofile-10)#quit
```

Step 2 Add an ONU to the OLT.

Connect an ONU to a GPON port of the OLT using a fiber. Services can be configured only after an ONU is successfully added to the OLT. Set the native VLAN of the port as the authentication VLAN.

```
huawei(config)#interface gpon 0/1
huawei(config-if-gpon-0/1)#ont add 0 0 sn-auth 1234567891234560 omci ont-lineprofile-id 10 ont-srvprofile-id 10
huawei(config-if-gpon-0/1)#ont port native-vlan 0 0 eth 1 vlan 100
```

Step 3 Create an S-VLAN and add an upstream port to it.

```
huawei(config)vlan 100 smart
huawei(config)port vlan 100 0/10 0
```

Step 4 Create service flows for 802.1X authentication.

```
huawei(config)#service-port 1 vlan 100 gpon 0/1/0 ont 0 gemport 0 multi-service user-vlan 100 tag-transform translate
```

Step 5 Configure the RADIUS protocol.

Set the IP address of the RADIUS server to 10.10.66.66 and set the shared key to 0123456789123456.

Configure a RADIUS server profile named far1 for 802.1X authentication.

```
huawei(config)#radius-server template far1
huawei(config-radius-far1)#radius-server shared-key-cipher 0123456789123456
huawei(config-radius-far1)#radius-server authentication 10.10.66.66 1812
huawei(config-radius-far1)#quit
```

Configure the RADIUS server profile named default for MAC bypass authentication, in which the MAC bypass authentication packet does not contain a domain name.

```
huawei(config)#radius-server template default
huawei(config-radius-default)#radius-server shared-key-cipher 0123456789123456
huawei(config-radius-default)#radius-server authentication 10.10.66.66 1812
huawei(config-radius-default)#undo radius-server user-name domain-included
huawei(config-radius-default)#quit
```

Configure the interface of the server source IP address for the OLT to communicate with the server.

```
huawei(config)#radius-server source-interface vlanif 999
```

Step 6 Configure the authentication scheme.

- Configure an 802.1X authentication scheme far1 and use RADIUS to authenticate users.

```
huawei(config)#aaa
huawei(config-aaa)#authentication-scheme far1
huawei(config-aaa-authen-far1)#authentication-mode radius
huawei(config-aaa-authen-far1)#quit
```

 Copyright © Huawei Technologies Co., Ltd.

- Create a domain named far1 and bind the created authentication scheme and RADIUS server profile to the domain.

  NOTE

  A domain is a group of users of the same type. Generally, in a user name in the format of userid@domain-name, *domain-name* following @ is the domain name, and *userid* is the user name used for identity authentication.

```
huawei(config-aaa)#domain far1
huawei(config-aaa-domain-far1)#authentication-scheme far1
huawei(config-aaa-domain-far1)#radius-server far1
```

- Configure a MAC bypass authentication scheme named default and use RADIUS to authenticate users.

```
huawei(config)#aaa
huawei(config-aaa)#authentication-scheme default
huawei(config-aaa-authen-default)#authentication-mode radius
huawei(config-aaa-authen-default)#quit
```

- Create a domain named default and bind the created authentication scheme and RADIUS server profile to the domain.

```
huawei(config-aaa)#domain default
huawei(config-aaa-domain-default)#authentication-scheme default
huawei(config-aaa-domain-default)#radius-server default
```

Step 7 Configure service port–based 802.1X authentication.

Enable the 802.1X global switch. Enable 802.1X authentication, MAC bypass authentication, and DHCP-triggered authentication on service port 1.

```
huawei(config)#dot1x enable
huawei(config)#dot1x service-port 1
huawei(config)#dot1x mac-bypass enable
huawei(config)#dot1x mac-bypass service-port 1
huawei(config)#dot1x dhcp-trigger enable
```

Step 8 Configure MAC address–based 802.1X authentication.

Enable the 802.1X global switch. Enable 802.1X authentication and DHCP-triggered authentication on service port 1.

NOTE

- The service VLAN ID for 802.1X authentication must be different from the guest VLAN ID. Otherwise, dynamic service port creation fails.
- If a service VLAN is configured for 802.1X authentication and the RADIUS authentication server delivers another service VLAN, the service VLAN delivered by the RADIUS authentication server is preferentially used.

```
huawei(config)#dot1x enable
huawei(config)#dot1x service-port auth-method 1 mac-based
huawei(config)#dot1x service-port 1
huawei(config)#dot1x service-vlan 20
huawei(config)#dot1x dhcp-trigger enable
```

----End

Operation Result

After the OLT is configured, it can be used as an access point for 802.1X authentication. A connected terminal can access the LAN by using the user name and password, or by using the MAC address.

 Copyright © Huawei Technologies Co., Ltd.

17.5.3 Core Switch Functioning as the 802.1X Authentication Device

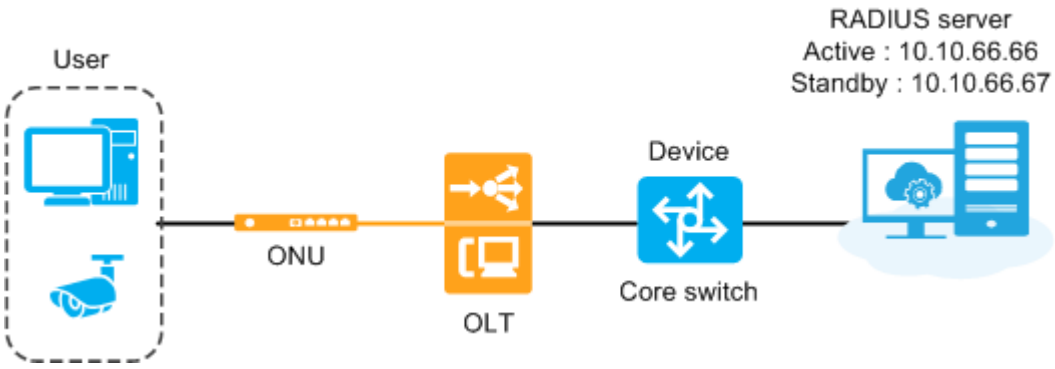
The core switch functions as the 802.1X authentication device, and the transparent transmission function is configured on the OLT for 802.1x protocol packets.

Prerequisites

- The ONU/OLT is online and in the normal state.
- The ONU has been configured to transparently transmit EAPOL and BPDU packets.

Figure 17-11 shows the networking diagram for RADIUS authentication.

Figure 17-11 Networking diagram for RADIUS authentication



Procedure

- Configure EAPOL packets to be transparently transmitted on the OLT.

```
huawei(config)#protocol permit-forwarding eapol enable
```

- Configure BPDU packets to be transparently transmitted on the OLT.

```
huawei(config)#vlan service-profile profile-id 10
huawei(config-vlan-srvprof-10)#bpdu tunnel enable
huawei(config-vlan-srvprof-10)#commit
huawei(config-vlan-srvprof-10)#quit
huawei(config)#vlan bind service-profile 11 profile-id 10
```

Operation Result

After the configuration is completed, 802.1x protocol packets can be transparently transmitted to the core switch, and the core switch then implements 802.1X authentication.

Configuration Script

```
protocol permit-forwarding eapol enable
vlan service-profile profile-id 10
bpdu tunnel enable
commit
quit
vlan bind service-profile 11 profile-id 10
```

 Copyright © Huawei Technologies Co., Ltd.

18 AAA

AAA refers to authentication, authorization, and accounting. In the process that a user accesses network resources, through AAA, certain rights are authorized to the user if the user passes authentication, and the original data about the user accessing network resources is recorded.

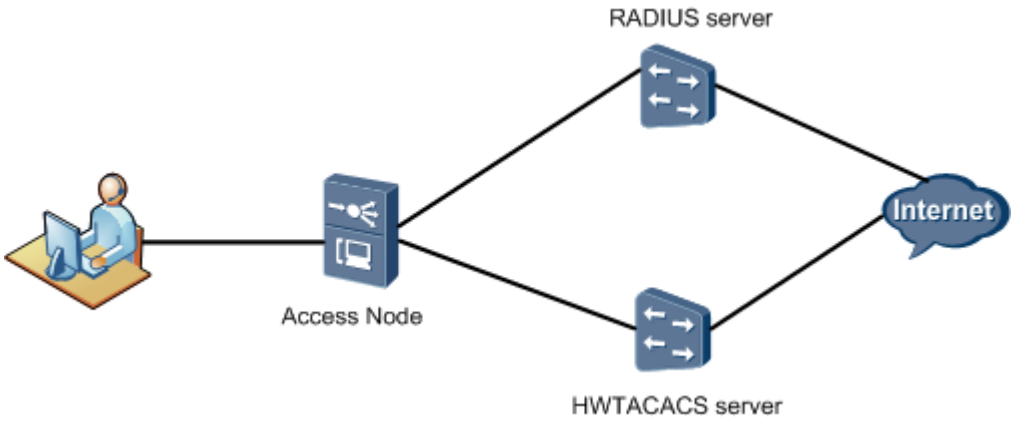
AAA includes three functions.

- Authentication: Checks whether a user is allowed to access network resources.
- Authorization: Determines what network resources a user can access.
- Accounting: Records the original data about the user accessing network resources.

AAA is generally applied to the users that access the Internet in the PPPoA, PPPoE, 802.1x, VLAN, WLAN or Admin SSH (associating the user name and the password with the domain name) mode.

In the existing network, 802.1x and Admin SSH correspond to the local AAA, that is, the access device functions as a local AAA server; PPPoE corresponds to the remote AAA, that is, the access device functions as the client of a remote AAA server.

The following picture shows an example network of the AAA application.



1. Authentication

   The MA5800 supports two authentication modes: local authentication, and remote authentication.

 Copyright © Huawei Technologies Co., Ltd.

- Local authentication: The user information (including the user name, password, and various attributes) is configured on the MA5800, and the MA5800 authenticates the user. This authentication mode is fast and can reduce carrier's cost; however, the amount of information that can be stored is limited by the device hardware.
- Remote authentication: The user information (including the user name, password, and various attributes of the user) is configured on an authentication server. The Remote Authentication Dial In User Service (RADIUS) protocol or HUAWEI Terminal Access Controller Access Control System (HWTACACS) protocol is used for remote authentication. The MA5800 serves as the authentication client and communicates with the RADIUS or HWTACACS server. When the RADIUS or HWTACACS server is faulty, the MA5800 can automatically switch to local authentication.

2. Authorization

   The MA5800 supports direct authorization, local authorization, HWTACACS authorization, and if-authenticated authorization.

   - Direct authorization: If trustful, a user can directly pass the authorization.
   - Local authorization: A user is locally authorized according to relevant attributes of the user configured on the MA5800.
   - HWTACACS authorization: The HWTACACS server authorizes a user.
   - If-authenticated authorization: If a user passes the authentication and the authentication mode is not non-authentication, the user passes the authorization.

3. Accounting

   The MA5800 supports non-accounting and remote accounting.

   - Non-accounting: A user is not charged.
   - Remote accounting: The MA5800 supports remote accounting through the AAA server.

The preceding figure shows that the AAA function can be implemented on the access device in the following three ways:

- The access device functions as a local AAA server. In this case, the local AAA needs to be configured. The local AAA does not support accounting.
- The access device functions as the client of a remote AAA server, and is connected to the RADIUS server through the RADIUS protocol, implementing the AAA. The RADIUS protocol, however, does not support authorization.
- The access device functions as the client of a remote AAA server, and is connected to the HWTACACS server through the HWTACACS protocol, implementing the AAA.

18.1 RADIUS

18.2 HWTACACS

18.3 Configuring the Local AAA
This topic describes how to configure the local AAA so that the user authentication can be performed locally.

18.4 Configuring the Remote AAA (RADIUS Protocol)
The MA5800 is interconnected with the RADIUS server through the RADIUS protocol to implement authentication and accounting.

 Copyright © Huawei Technologies Co., Ltd.

18.5 Configuration Example of the RADIUS Authentication and Accounting
The MA5800 is interconnected with the RADIUS server through the RADIUS protocol to implement authentication and accounting.

18.6 Configuring the Remote AAA (HWTACACS Protocol)
The MA5800 is interconnected with the HWTACACS server through the HWTACACS protocol to implement authentication, authorization, and accounting.

18.7 Configuration Example of the RADIUS AAA (802.1X access user)
The MA5800 is interconnected with the RADIUS server through the RADIUS protocol to implement authentication.

18.8 Configuration Example of HWTACACS Authentication (Management User)
The MA5800 allows the management user of the device to log in to the system in the HWTACACS authentication mode.

18.1 RADIUS

Definition

- RADIUS is short for the remote authentication dial-in user service. It is a distributed information interaction protocol with the client-server structure. Generally, it is used to manage a large number of distributed dial-in users.
- RADIUS implements the user accounting by managing a simple user authentication database.
- The authentication and accounting requests of users can be passed on to the RADIUS server through a network access server (NAS).

Principle

- When a user tries to access another network (or some network resources) by setting up a connection to the NAS through a network, the NAS forwards the user authentication and accounting information to the RADIUS server. The RADIUS protocol specifies the means of transmitting the user information and accounting information between the NAS and the RADIUS server.
- The RADIUS server receives the connection requests of users sent from the NAS, authenticates the user account and password contained in the user authentication data, and returns the required data to the NAS.

Message Flow of the RADIUS Protocol

The RADIUS server stores the user names and passwords in a unique user authentication database for authenticating the users. When a user wishes to connect to an NE through a device and then obtain the right to access the Internet or access certain network resources, the NE authenticates the user or the corresponding connection.

The NE sends the authentication, authorization, and accounting information of the user to the RADIUS server. The RADIUS protocol specifies how the NE and the RADIUS server should exchange the user information and the accounting information. The RADIUS server receives the connection request of the user, authenticates the user, and sends the necessary configuration information of the user to the NE. The exchange of authentication information between the NE and

 Copyright © Huawei Technologies Co., Ltd.

the RADIUS server is key protected. This protects the user password against any interception when the password is transmitted over an insecure network. **Figure 18-1** shows the message flow between the RADIUS client and the RADIUS server.

Figure 18-1 Message flow between the RADIUS client and the RADIUS server

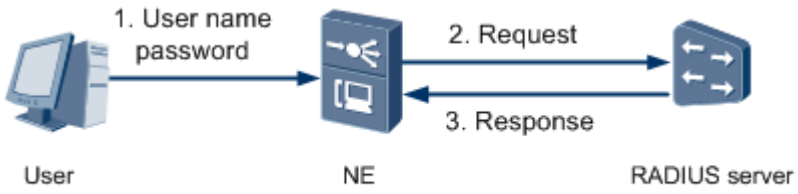
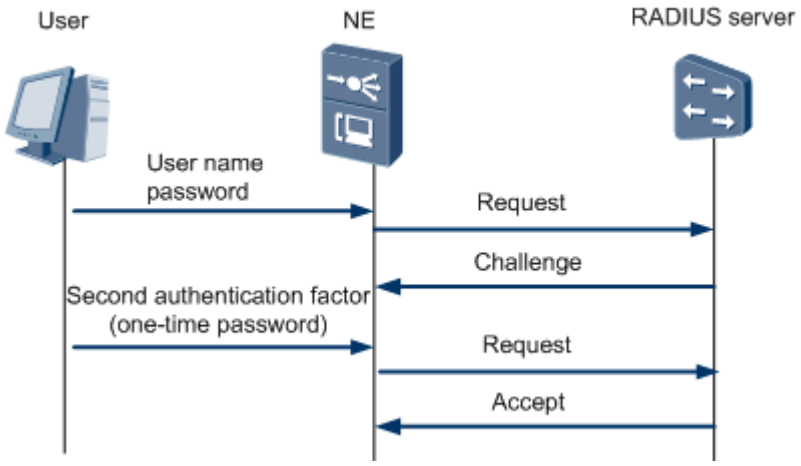


Figure 18-2 Two-factor authentication flow between the RADIUS client and the RADIUS server



NOTE

An NE refers to an access device that can function as a RADIUS client.

1. When a user logs in to the NE, the corresponding user name, password, and one-time password (required for two-factor authentication) are sent to the NE.
2. The RADIUS client on the NE receives the user name and password, and sends an authentication request to the RADIUS server.
3. The RADIUS server receives the legal request, authenticates the user, and sends the necessary authorization information of the user to the RADIUS client.

The authentication information exchanged between the RADIUS client and the RADIUS server must be encrypted before being transmitted over the network. Otherwise, the information may be intercepted when the network is insecure.

RADIUS supports the management VPN: When a carrier uses VPN to isolate management and service channels, a separated VPN can be used through RADIUS

 Copyright © Huawei Technologies Co., Ltd.

to manage user authentication. When such a VPN is used, only the VPN used by RADIUS needs to be specified.

The accounting message flow is similar to the authentication/authorization message flow.

18.2 HWTACACS

Definition

HWTACACS (HUAWEI Terminal Access Controller Access Control System) is a security protocol enhanced based on TACACS+(draft-grant-tacacs-02). It is compatible with TACACS+ but not compatible with TACACS. Similar to the RADIUS protocol, HWTACACS implements AAA functions for multiple users by communicating with the HWTACACS server in the client/server (C/S) mode.

Purpose

HWTACACS is used for the authentication, authorization, and accounting of access users and administrators.

Principle

Adopting the client/server architecture, HWTACACS is a protocol through which the network access server (NAS) (MA5800) transmits the encrypted HWTACACS data packets to communicate with the HWTACACS database of the security server. The working mode is as follows:

- HWTACACS authentication. After being set up a connection with the NAS port, the NAS communicates with the daemon of the HWTACACS server. After the user input the user name, password, and dynamic password (if required) according to the prompt messages, the NAS sends the information to the daemon.
- HWTACACS authorization. After being authenticated, the user can be authorized. The NAS communicates with the daemon of the HWTACACS server, and then returns the accept or reject response of the authorization.

Differences Between HWTACACS and RADIUS

The HWTACACS message flow is similar to the RADIUS message flow. The difference is that, in the HWTACACS message flow, the server returns an authentication response rather than the user right after the user passes authentication. The user right is returned only when the authorization process is completed.

HWTACACS features more reliable transmission and encryption than RADIUS and is more suitable for security control. The following table lists the differences between HWTACACS and RADIUS.

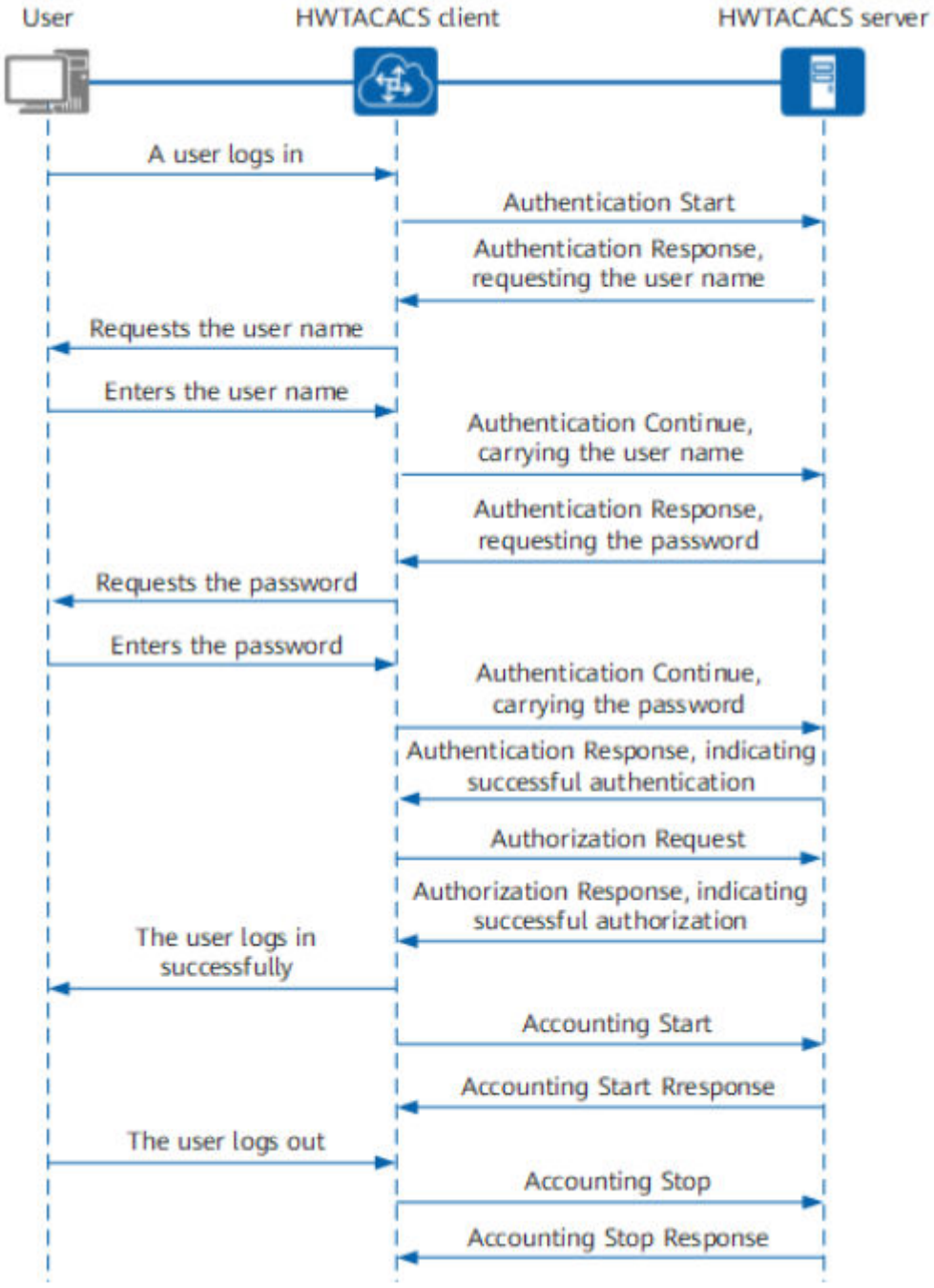
 Copyright © Huawei Technologies Co., Ltd.

HWTACACS	RADIUS
Uses TCP to ensure more reliable network transmission.	Uses UDP for transmission.
Encrypts the body of HWTACACS packets, except their header.	Encrypts only the password field of the authenticated packets.
Separated authorization and authentication.	Concurrent processing of authentication and authorization.
Applicable to security control.	Applicable to accounting.
Supports authorization of the configuration commands on the router.	Does not support the authorization of the configuration commands on the router.

HWTACACS Interaction Process

This section describes how HWTACACS performs authentication, authorization, and accounting for Telnet users. **Figure 18-3** shows the message exchange process.

Figure 18-3 HWTACACS message interaction



The following describes the HWTACACS message exchange process shown in **Figure 18-3**:

1. A Telnet user sends a request packet.
2. After receiving the request packet, the HWTACACS client sends an Authentication Start packet to the HWTACACS server.
3. The HWTACACS server sends an Authentication Response packet to request the user name.

 Copyright © Huawei Technologies Co., Ltd.

4. After receiving the Authentication Response packet, the HWTACACS client sends a packet to query the user name.
5. The user enters the user name.
6. The HWTACACS client sends an Authentication Continue packet containing the user name to the HWTACACS server.
7. The HWTACACS server sends an Authentication Response packet to request the password.
8. After receiving the Authentication Response packet, the HWTACACS client queries the password.
9. The user enters the password.
10. The HWTACACS client sends an Authentication Continue packet containing the password to the HWTACACS server.
11. The HWTACACS server sends an Authentication Response packet, indicating that the user has been authenticated.
12. The HWTACACS client sends an Authorization Request packet to the HWTACACS server.
13. The HWTACACS server sends an Authorization Response packet, indicating that the user has been authorized.
14. The HWTACACS client receives the Authorization Response packet and displays the login page.
15. The HWTACACS client sends an Accounting Request (start) packet to the HWTACACS server.
16. The HWTACACS server sends an Accounting Response packet.
17. The user requests to go offline.
18. The HWTACACS client sends an Accounting Request (stop) packet to the HWTACACS server.
19. The HWTACACS server sends an Accounting Response packet.

18.3 Configuring the Local AAA

This topic describes how to configure the local AAA so that the user authentication can be performed locally.

Context

- The local AAA configuration is simple, which does not depend on the external server.
- The local AAA supports authentication and authorization.

Procedure

Step 1 Configure the AAA authentication scheme.

NOTE

- The authentication scheme specifies how all the users in an Internet service provider (ISP) domain are authenticated. The system supports up to 32 authentication schemes.
- The system has a default authentication scheme named **default**. It can be modified, but cannot be deleted.

 Copyright © Huawei Technologies Co., Ltd.

1. Run the **aaa** command to enter the AAA mode.
2. Run the **authentication-scheme** command to add an authentication scheme.
3. Run the **authentication-mode local** command to configure the authentication mode of the authentication scheme.
4. Run the **quit** command to return to the AAA mode.

Step 2 Create a domain.

NOTE

- A domain is a group of users of the same type.
- In the user name format userid@domain-name (for example, userdomain@domain.net), "userid" indicates the user name for authentication and "domain-name" followed by "@" indicates the domain name.
- The domain name for user login cannot exceed 20 characters.
- For a user with a domain name, the total length of the user name plus @ plus the domain name cannot exceed 31 characters.

In the AAA mode, run the **domain** command to create a domain.

Step 3 Refer the authentication scheme.

NOTE

You can use an authentication scheme in a domain only after the authentication scheme is created.

1. In the domain mode, run the **authentication-scheme** command to use the authentication scheme.
2. Run the **quit** command to return to the AAA mode.

Step 4 Create a local AAA user.

In the AAA mode, run the **local-user password** command to create a local AAA user.

Step 5 Set the access type of a local AAA user.

In the AAA mode, run the **local-user service-type** command to set the access type of a local user.

Step 6 Set the access priority of a local AAA user.

In the AAA mode, run the **local-user privilege level** command to set the access priority of a local user.

----End

Example

User 1 in the ISP domain adopts the local server for authentication. The authentication scheme is newscheme, the password is configured, and access priority is 1, do as follows:

```
huawei(config)#aaa
huawei(config-aaa)#authentication-scheme newscheme
  Info: Create a new authentication scheme.
huawei(config-aaa-authen-newscheme)#authentication-mode local
huawei(config-aaa-authen-newscheme)#quit
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-aaa)#domain isp
  Info: Create a new domain.
huawei(config-aaa-domain-isp)#authentication-scheme newscheme
huawei(config-aaa-domain-isp)#quit
huawei(config-aaa)#local-user user1@isp password
huawei(config-aaa)#local-user user1@isp service-type terminal
huawei(config-aaa)#local-user user1@isp privilege level 1
```

18.4 Configuring the Remote AAA (RADIUS Protocol)

The MA5800 is interconnected with the RADIUS server through the RADIUS protocol to implement authentication and accounting.

Context

- What is RADIUS:
  - RADIUS is short for the remote authentication dial-in user service. It is a distributed information interaction protocol with the client-server structure. Generally, it is used to manage a large number of distributed dial-in users.
  - RADIUS implements the user accounting by managing a simple user authentication database.
  - The authentication and accounting requests of users can be passed on to the RADIUS server through a network access server (NAS).
- Principle of RADIUS:
  - When a user tries to access another network (or some network resources) by setting up a connection to the NAS through a network, the NAS forwards the user authentication and accounting information to the RADIUS server. The RADIUS protocol specifies the means of transmitting the user information and accounting information between the NAS and the RADIUS server.
  - The RADIUS server receives the connection requests of users sent from the NAS, authenticates the user account and password contained in the user authentication data, and returns the required data to the NAS.
- Specification:
  - For the MA5800, the RADIUS is configured based on each RADIUS server group.
  - In actual networking, a RADIUS server group can be any of the following:
    - An independent RADIUS server
    - A pair of primary/secondary RADIUS servers with the same configuration but different IP addresses
  - The following lists the attributes of a RADIUS server template:
    - Primary and secondary authentication servers
    - Primary and secondary accounting servers
    - Shared key
    - Retransmission parameter

 Copyright © Huawei Technologies Co., Ltd.

  - Timeout parameter
  - Whether a user name contains the domain name
- The configuration of the RADIUS protocol defines only the essential parameters for the information exchange between the MA5800 and the RADIUS server. To make the essential parameters take effect, the RADIUS server group should be referenced in a certain domain.
- The RADIUS attribute list defines the attribute parameters for interaction between the MA5800 and the RADIUS server. **Table 18-1** describes the parameters.

Table 18-1 RADIUS attribute list

Parameter Code	Parameter Name	Description
1	User-Name	Indicates the user name for authentication.
2	Password	Indicates the user password for authentication. This parameter is valid only for PAP authentication.
6	Service-Type	Indicates the user service type. The value of this parameter is 2 (frame) for access users and is 6 for remote management users.
7	Framed-Protocol	The value of this parameter is fixed to 7 (PPP) because ITU-T RFC2865 does not define 802.1x for this parameter.
24	State	If the access challenge packet that the RADIUS server sends to a device contains this parameter, the subsequent access request packet sent by the device to the RADIUS server must also contain this parameter of the same value as that is contained in the access challenge packet.

 Copyright © Huawei Technologies Co., Ltd.

Parameter Code	Parameter Name	Description
25	Class	If the access accept packet sent by the RADIUS server to a device contains this parameter, the subsequent charging request packet sent by the device to the RADIUS server must also contain this parameter of the same value. For a standard RADIUS server, a device can use the Class attribute to represent the CAR parameter.
27	Session-Timeout	Indicates the available remaining time in the unit of second. It is the user re-authentication time in the EAP challenge packet.
29	Termination-Action	Indicates the service termination mode. The valid modes are re-authentication and forcing users to go offline.
32	NAS-Identifier	Indicates the host name of the device.
40	Acct-Status-Type	Indicates the charging packet type. • 1: charging start packet • 2: charging stop packet • 3: real-time charging packet
41	Acct-Delay-Time	Indicates the time for generating a charging packet in the unit of second.
44	Acct-Session-Id	Indicates the charging connection number. The connection numbers for the charging start packet, real-time charging packet, and charging stop packet of the same connection must be the same.

 Copyright © Huawei Technologies Co., Ltd.

Parameter Code	Parameter Name	Description
45	Acct-Authentic	Indicates the user authentication mode. ● 1: RADIUS authentication ● 2: local authentication
46	Acct-Session-Time	Indicates the time for a user to go online in the unit of second.
49	Terminate-Cause	Indicates the user connection interruption cause. The valid values are as follows: ● User-Request(1): The user actively goes offline. ● Lost Carrier(2): The handshake fails, such as the EAPOL detection fails. ● User Error(17): The user authentication fails or times out.
52	Acct-Input-Gigawords	Indicates the number of upstream bytes. The unit can be 4GByte, kbyte, Mbyte, or Gbyte, depending on the command configuration.
53	Acct-Output-Gigawords	Indicates the number of downstream bytes. The unit can be 4GByte, kbyte, Mbyte, or Gbyte, depending on the command configuration.
55	Event-Timestamp	Indicates the user online time in the unit of second. The value is the absolute number of seconds counting from 1970-01-01 00:00:00.
61	NAS-Port-Type	Indicates the NAS port type.
80	Message-Authenticator	Verifies validity of packets between the RADIUS server and RADIUS client to prevent malicious attacks.

 Copyright © Huawei Technologies Co., Ltd.

Parameter Code	Parameter Name	Description
26-29 **NOTE** The preceding parameters are RADIUS standard attributes. Starting from this row, the following parameters are Huawei-defined attributes.	Exec-Privilege	Indicates the priority of operation users such as SSH users. The value ranges from 0 to 15. • 0: common user • 1: operator • 2-15: administrator • >15: common user
26-254	Version	Indicates the software version of the access device.
26-255	Product-ID	Indicates the product name.

Procedure

Step 1 Configure the authentication scheme.

NOTE

- The authentication scheme specifies how all the users in an ISP domain are authenticated.
- The system supports up to 32 authentication schemes. The system has a default accounting scheme named **default**. It can only be modified, but cannot be deleted.

1. Run the **aaa** command to enter the AAA mode.
2. Run the **authentication-scheme** command to add an authentication scheme.
3. Run the **authentication-mode radius** command to configure the authentication mode of the authentication scheme.
4. Run the **quit** command to return to the AAA mode.

Step 2 Configure the accounting scheme.

NOTE

- The accounting scheme specifies how all the users in an ISP domain are charged.
- The system supports up to 256 accounting schemes. The system has a default accounting scheme named **default**. It can be modified, but cannot be deleted.

1. In the AAA mode, run the **accounting-scheme** command to add an AAA accounting scheme.
2. Run the **accounting-mode radius** command to configure the accounting mode.
3. Run the **quit** command to return to the AAA mode.
4. Run the **quit** command to return to the global config mode.

Step 3 Configure the RADIUS server template.

1. Run the **radius-server template** command to create an RADIUS server template and enter the RADIUS server template mode.

 Copyright © Huawei Technologies Co., Ltd.

2. Run the **radius-server authentication** command to configure the IP address and the UDP port ID of the RADIUS server for authentication.

   NOTE

   - To guarantee normal communication between the MA5800 and the RADIUS server, before configuring the IP address and UDP port of the RADIUS server, make sure that the route between the RADIUS server and the MA5800 is in the normal state.
   - Make sure that the configuration of the RADIUS service port of the MA5800 is consistent with the port configuration of the RADIUS server.

3. Run the **radius-server accounting** command to configure the IP address and the UDP port ID of the RADIUS server for accounting.
4. (Optional) Run the **radius-server shared-key** command to configure the shared key of the RADIUS server.

   NOTE

   - The RADIUS client (MA5800) and the RADIUS server use the MD5 algorithm to encrypt the RADIUS packets. They check the validity of the packets by setting the encryption key. They can receive the packets from each other and can respond to each other only when their keys are the same.
   - You are advised to periodically update keys to ensure key security.

5. (Optional) Run the **radius-server timeout** command to set the response timeout time of the RADIUS server. By default, the timeout time is 5s.

   The MA5800 sends the request packets to the RADIUS server. If the RADIUS server does not respond within the response timeout time, it re-transmits the request packets to the RADIUS to ensure that users can get corresponding services from the RADIUS server.

6. (Optional) Run the **radius-server retransmit** command to set the maximum re-transmit time of the RADIUS request packets. By default, the maximum re-transmit time is 3.

   When the re-transmit time of the RADIUS request packets from the MA5800 to a RADIUS server exceeds the maximum re-transmit time, the MA5800 considers that its communication with theRADIUS server is interrupted (optional), and therefore transmits the RADIUS request packets to another RADIUS server.

7. (Optional) Run the **(undo)radius-server user-name domain-included** command to configure the user name (not) to carry the domain name when transmitted to the RADIUS server. By default, the user name of the RADIUS server carries the domain name.
   - An access user is named in the format of **userid@domain-name**, and the part after @ is the domain name. The MA5800 classifies a user into a domain according to the domain name.
   - If an RADIUS server group rejects the user name carrying the domain name, the RADIUS server group cannot be set or used in two or more domains. Otherwise, when some access users in different domains have the same user name, the RADIUS server considers that these users are the same because the names transmitted to the server are the same.
8. Run the **quit** command to return to the global config mode.

Step 4 Create a domain.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

- A domain is a group of users of the same type.
- In the user name format userid@domain-name (for example, userdomain@domain.net), "userid" indicates the user name for authentication and "domain-name" followed by "@" indicates the domain name.
- The domain name for user login cannot exceed 20 characters.
- For a user with a domain name, the total length of the user name plus @ plus the domain name cannot exceed 31 characters.

1. Run the **aaa** command to enter the AAA mode.
2. In the AAA mode, run the **domain** command to create a domain.

Step 5 Use the authentication scheme.

You can use an authentication scheme in a domain only after the authentication scheme is created.

In the domain mode, run the **authentication-scheme** command to use the authentication scheme.

Step 6 Use the accounting scheme.

You can use an accounting scheme in a domain only after the accounting scheme is created.

In the domain mode, run the **accounting-scheme** command to use the accounting scheme.

Step 7 Use the RADIUS server template.

NOTE

You can use a RADIUS server template in a domain only after the RADIUS server template is created.

1. In the domain mode, run the **radius-server** command to use the RADIUS server template.
2. Run the **quit** command to return to the AAA mode.

Step 8 Run the terminal **user authentication-mode AAA** command to configure automatically add a domain name to the user name for authentication when a management user logs in.

----End

Example

User 1 in the ISP domain adopts the HWTACACS protocol for authentication and accounting. The RADIUS server with the IP address 10.10.66.66 functions as the primary server for authentication and accounting. The server with the IP address 10.10.66.67 functions as the secondary server for authentication and accounting. The authentication port number is 1812, the accounting port number is 1813, and set the shared key. Other parameters adopt the default settings.

```
huawei(config)#aaa
huawei(config-aaa)#authentication-scheme newscheme
huawei(config-aaa-authen-newscheme)#authentication-mode radius
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-aaa-authen-newscheme)#quit
huawei(config-aaa)#accounting-scheme newscheme
huawei(config-aaa-accounting-newscheme)#accounting-mode radius
huawei(config-aaa-accounting-newscheme)#quit
huawei(config-aaa)#quit
huawei(config)#radius-server template hwtest
huawei(config-radius-hwtest)#radius-server shared-key
huawei(config-radius-hwtest)#radius-server authentication 10.10.66.66 1812
huawei(config-radius-hwtest)#radius-server authentication 10.10.66.67 1812 secondary
huawei(config-radius-hwtest)#radius-server accounting 10.10.66.66 1813
huawei(config-radius-hwtest)#radius-server accounting 10.10.66.67 1813 secondary
huawei(config-radius-hwtest)#quit
huawei(config)#aaa
huawei(config-aaa)#domain isp
huawei(config-aaa-domain-isp)#authentication-scheme newscheme
huawei(config-aaa-domain-isp)#accounting-scheme newscheme
huawei(config-aaa-domain-isp)#radius-server hwtest
huawei(config-aaa-domain-isp)#quit
huawei(config-aaa)#quit
huawei(config)#terminal user authentication-mode AAA isp
```

18.5 Configuration Example of the RADIUS Authentication and Accounting

The MA5800 is interconnected with the RADIUS server through the RADIUS protocol to implement authentication and accounting.

Service Requirements

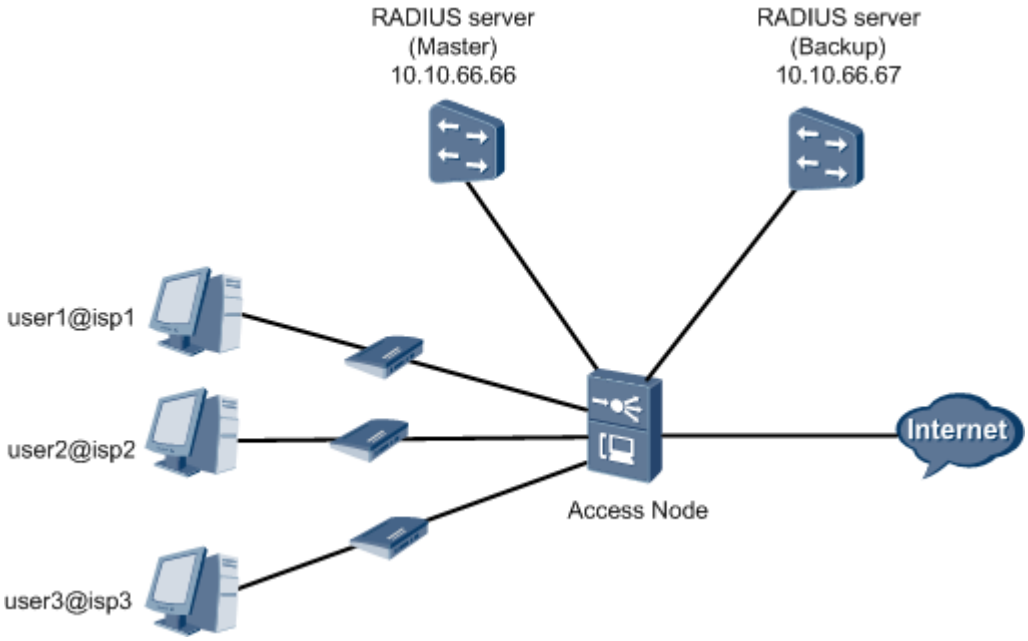
- The RADIUS server performs authentication and accounting for User 1 in the isp1 domain.
- The RADIUS server with the IP address 10.10.66.66 functions as the primary server for authentication and accounting.
- The RADIUS server with the IP address 10.10.66.67 functions as the secondary server for authentication and accounting.
- The authentication port number is 1812, and the accounting port number is 1813.
- Set shared key of the RADIUS server.
- Other parameters adopt the default settings.

Networking

Figure 18-4 shows an example network of the RADIUS Authentication and Accounting application.

 Copyright © Huawei Technologies Co., Ltd.

Figure 18-4 Example network of the RADIUS Authentication and Accounting application.



Procedure

Step 1 Configure the authentication scheme.

Configure authentication scheme named **newscheme** (users are authenticated through RADIUS).

```
huawei(config)#aaa
huawei(config-aaa)#authentication-scheme newscheme
huawei(config-aaa-authen-newscheme)#authentication-mode radius
huawei(config-aaa-authen-newscheme)#quit
```

Step 2 Configure the accounting scheme.

Configure accounting scheme named **newscheme** (users are accounted through RADIUS).

```
huawei(config-aaa)#accounting-scheme newscheme
huawei(config-aaa-accounting-newscheme)#accounting-mode radius
huawei(config-aaa-accounting-newscheme)#quit
huawei(config-aaa)#quit
```

Step 3 Configure the RADIUS protocol.

Create RADIUS server template named **hwtest** with the RADIUS server 10.10.66.66 as the primary authentication and accounting server, the RADIUS server 10.10.66.67 as the secondary authentication and accounting server, and set the shared key.

```
huawei(config)#radius-server template hwtest
  Info: A new server-template is created.
  Warning: Please configure the shared-key. Configuring shared-key is mandatory
to communicate with RADIUS server.
huawei(config-radius-hwtest)#radius-server shared-key
huawei(config-radius-hwtest)#radius-server authentication 10.10.66.66 1812
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-radius-hwtest)#radius-server authentication 10.10.66.67 1812 secondary
huawei(config-radius-hwtest)#radius-server accounting 10.10.66.66 1813
huawei(config-radius-hwtest)#radius-server accounting 10.10.66.67 1813 secondary
huawei(config-radius-hwtest)#quit
```

Step 4 Create a domain.

Create a domain named isp1.

```
huawei(config)#aaa
huawei(config-aaa)#domain isp1
  Info: Create a new domain.
```

Step 5 Use the authentication scheme.

You can use an authentication scheme in a domain only after the authentication scheme is created.

```
huawei(config-aaa-domain-isp1)#authentication-scheme newscheme
```

Step 6 Use the accounting scheme.

You can use an accounting scheme in a domain only after the accounting scheme is created.

```
huawei(config-aaa-domain-isp1)#accounting-scheme newscheme
```

Step 7 Use the RADIUS server template.

You can use a RADIUS server template in a domain only after the RADIUS server template is created.

```
huawei(config-aaa-domain-isp1)#radius-server hwtest
huawei(config-aaa-domain-isp1)#quit
```

----End

Result

User 1 in isp1 can pass authentication only if both the user name and password are correct, and then can log in to theMA5800. Then, the user starts to be accounted.

Configuration File

```
aaa
authentication-scheme newscheme
authentication-mode radius
quit
accounting-scheme newscheme
accounting-mode radius
quit
quit
radius-server template hwtest
radius-server shared-key
radius-server authentication 10.10.66.66 1812
radius-server authentication 10.10.66.67 1812 secondary
radius-server accounting 10.10.66.66 1813
radius-server accounting 10.10.66.67 1813 secondary
quit
aaa  domain isp1 authentication-scheme newscheme
accounting-scheme newscheme
radius-server hwtest
quit
```

 Copyright © Huawei Technologies Co., Ltd.

18.6 Configuring the Remote AAA (HWTACACS Protocol)

The MA5800 is interconnected with the HWTACACS server through the HWTACACS protocol to implement authentication, authorization, and accounting.

Context

- What is HWTACACS:
  - HWTACACS is a security protocol with enhanced functions on the base of TACACS+(draft-grant-tacacs-02). Similar to the RADIUS protocol, HWTACACS implements multiple subscriber AAA functions through communications with the HWTACACS server in the client/server (C/S) mode.
  - HWTACACS is used for the authentication, authorization, and accounting for the management users.
- Principle of HWTACACS:

  Adopting the client/server architecture, HWTACACS is a protocol through which the NAS (MA5800) transmits the encrypted HWTACACS data packets to communicate with the HWTACACS database of the security server. The working mode is as follows:
  - HWTACACS authentication. When the remote user connects to the corresponding port of the NAS, the NAS communicates with the daemon of the HWTACACS server, and obtains the prompt of entering the user name from the daemon. Then, the NAS displays the message to the user. When the remote user enters the user name, the NAS transmits the user name to the daemon. Then, the NAS obtains the prompt of entering the password, and displays the message to the user. After the remote user enters the password, the NAS transmits the password to the daemon.
  - HWTACACS authorization. After being authenticated, the user can be authorized. The NAS communicates with the daemon of the HWTACACS server, and then returns the accept or reject response of the authorization.

NOTE

- The HWTACACS configuration only defines the parameters used for data exchange between the MA5800 and the HWTACACS server. To make these parameters take effect, you need to use the HWTACACS server group in a domain.
- The settings of an HWTACACS server template can be modified regardless of whether the template is bound to a server or not.

Procedure

Step 1 Configure the AAA authentication scheme.

The authentication scheme specifies how all the users in an ISP domain are authenticated.

The system supports up to 32 authentication schemes. The system has a default authentication scheme named **default**. It can be modified, but cannot be deleted.

 Copyright © Huawei Technologies Co., Ltd.

1. Run the **aaa** command to enter the AAA mode.
2. Run the **authentication-scheme** command to add an authentication scheme.
3. Run the **authentication-mode hwtacacs** command to configure the authentication mode of the authentication scheme. Use the HWTACACS protocol to authenticate users.
4. Run the **quit** command to return to the AAA mode.

Step 2 Configure the AAA authorization scheme.

The authorization scheme specifies how all the users in an ISP domain are authorized.

1. In the AAA mode, run the **authorization-scheme** command to add an AAA authorization scheme.
2. Run the **authorization-mode hwtacacs** command to configure the authorization mode.
3. Run the **quit** command to return to the AAA mode.

Step 3 Configure the AAA accounting scheme.

The accounting scheme specifies how all the users in an ISP domain are charged.

The system supports up to 256 accounting schemes. The system has a default accounting scheme named **default**. It can be modified, but cannot be deleted.

1. In the AAA mode, run the **accounting-scheme** command to add an AAA accounting scheme.
2. Run the **accounting-mode hwtacacs** command to configure the accounting mode. By default, the accounting is not performed.
3. Run the **quit** command to return to the AAA mode.
4. Run the **quit** command to return to the global config mode.

Step 4 Configure the HWTACACS protocol.

The configuration of the HWTACACS protocol of the MA5800 is on the basis of the HWTACACS server group. In actual networking scenarios, an HWTACACS server group can be an independent HWTACACS server or a combination of two HWTACACS servers, that is, a primary server and a secondary server with the same configuration but different IP addresses.

Each HWTACACS server template contains the primary/secondary server IP address, shared key, and HWTACACS server type.

Primary and secondary authentication, accounting, and authorization servers can be configured. The IP address of the primary server, however, must be different from that of the secondary server. Otherwise, the configuration of primary and secondary servers will fail. By default, the IP addresses of the primary and secondary servers are both 0.0.0.0.

1. Run the **hwtacacs-server template** command to create an HWTACACS server template and enter the HWTACACS server template mode.
2. Run the **hwtacacs-server authentication** command to configure a primary authentication server. You can select **secondary** to configure a secondary authentication server.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

- To ensure normal communication between the MA5800 and the HWTACACS server, before configuring the IP address and the TCP port of the HWTACACS server, make sure that the route between the HWTACACS server and the MA5800 is in the normal state.
- Make sure that the HWTACACS server port of the MA5800 is the same as the port of the HWTACACS server.

3. Run the **hwtacacs-server accounting** command to configure a primary accounting server. You can select **secondary** to configure a secondary accounting server.
4. Run the **hwtacacs-server authorization** command to configure a primary authorization server. You can select **secondary** to configure a secondary authorization server.
5. (Optional) Run the **hwtacacs-server shared-key** command to configure the shared key of the HWTACACS server.

   NOTE

   - The HWTACACS client (MA5800) and the HWTACACS server use the MD5 algorithm to encrypt the HWTACACS packets. They check the validity of the packets by configuring the encryption key. They can receive the packets from each other and can respond to each other only when their keys are the same.
   - By default, the HWTACACS server does not have a key. You are advised to configure and periodically update keys to ensure key security.

6. (Optional) Run the **hwtacacs-server timer response-timeout** to set the response timeout time of the HWTACACS server.

   NOTE

   - If the HWTACACS server does not respond to the HWTACACS request packets within the timeout time, the communication with the current HWTACACS server is considered as interrupted.
   - By default, the response timeout time of the HWTACACS server is 5s.

7. (Optional) In the global config mode, run the **hwtacacs-server accounting-stop-packet** command to configure the re-transmission mechanism of the accounting-stop packets of the HWTACACS server.

   NOTE

   - To prevent the loss of the accounting packets, the MA5800 supports the re-transmission of the accounting-stop packets of the HWTACACS server.
   - By default, the re-transmit time of the accounting-stop packets of the HWTACACS server is 100.

8. (Optional) Run the **(undo) hwtacacs-server user-name domain-included** command to configure the user name (not) to carry the domain name when transmitted to the HWTACACS server.
   - By default, the user name of the HWTACACS server carries the domain name.
   - After the **undo hwtacacs-server user-name domain-included** command is executed, the domain name is deleted from the user name when the client sends authentication and authorization requests to the HWTACACS server. The domain name in the user name of the accounting request is, however, reserved. This is to ensure that the users can be distinguished from each other in the accounting.

 Copyright © Huawei Technologies Co., Ltd.

9. Run the **quit** command to return to the global config mode.

Step 5 Create a domain.

NOTE

- A domain is a group of users of the same type.
- In the user name format userid@domain-name (for example, userdomain@domain.net), "userid" indicates the user name for authentication and "domain-name" followed by "@" indicates the domain name.
- The domain name for user login cannot exceed 20 characters.
- For a user with a domain name, the total length of the user name plus @ plus the domain name cannot exceed 31 characters.

1. Run the **aaa** command to enter the AAA mode.
2. In the AAA mode, run the **domain** command to create a domain.

Step 6 Use the authentication scheme.

You can use an authentication scheme in a domain only after the authentication scheme is created.

In the domain mode, run the **authentication-scheme** command to use the authentication scheme.

Step 7 Use the accounting scheme.

You can use an accounting scheme in a domain only after the accounting scheme is created.

In the domain mode, run the **accounting-scheme** command to use the accounting scheme.

Step 8 Use the authorization scheme.

You can use an authorization scheme in a domain only after the authorization scheme is created.

In the domain mode, run the **authorization-scheme** command to use the authorization scheme.

Step 9 Use the HWTACACS server template.

You can use an HWTACACS server template in a domain only after the HWTACACS server template is created.

1. In the domain mode, run the **hwtacacs-server** command to use the HWTACACS server template.
2. Run the **quit** command to return to the AAA mode.

Step 10 Run the terminal **user authentication-mode AAA** command to configure automatically add a domain name to the user name for authentication when a management user logs in.

----End

Example

User 1 in the ISP domain adopts the HWTACACS protocol for authentication, authorization, and accounting. HWTACACS server 10.10.66.66 functions as the

 Copyright © Huawei Technologies Co., Ltd.

primary authentication, authorization, and accounting server, and HWTACACS server 10.10.66.67 functions as the standby authentication, authorization, and accounting server. On the HWTACACS server, set the shared key, the parameters adopt the default values. To perform the preceding configuration, do as follows:

```
huawei(config)#aaa
huawei(config-aaa)#authentication-scheme newscheme
huawei(config-aaa-authen-newscheme)#authentication-mode hwtacacs
huawei(config-aaa-authen-newscheme)#quit
huawei(config-aaa)#authorization-scheme newscheme
huawei(config-aaa-author-newscheme)#authorization-mode hwtacacs
huawei(config-aaa-author-newscheme)#quit
huawei(config-aaa)#accounting-scheme newscheme
huawei(config-aaa-accounting-newscheme)#accounting-mode hwtacacs
huawei(config-aaa-accounting-newscheme)#quit
huawei(config-aaa)#quit
huawei(config)#hwtacacs-server template hwtest
huawei(config-hwtacacs-hwtest)#hwtacacs-server shared-key
huawei(config-hwtacacs-hwtest)#hwtacacs-server authentication 10.10.66.66
huawei(config-hwtacacs-hwtest)#hwtacacs-server authentication 10.10.66.67 secondary
huawei(config-hwtacacs-hwtest)#hwtacacs-server authorization 10.10.66.66
huawei(config-hwtacacs-hwtest)#hwtacacs-server authorization 10.10.66.67 secondary
huawei(config-hwtacacs-hwtest)#hwtacacs-server accounting 10.10.66.66
huawei(config-hwtacacs-hwtest)#hwtacacs-server accounting 10.10.66.67 secondary
huawei(config-hwtacacs-hwtest)#quit
huawei(config)#aaa
huawei(config-aaa)#domain isp
huawei(config-aaa-domain-isp)#authentication-scheme newscheme
huawei(config-aaa-domain-isp)#authorization-scheme newscheme
huawei(config-aaa-domain-isp)#accounting-scheme newscheme
huawei(config-aaa-domain-isp)#hwtacacs-server hwtest
huawei(config-aaa-domain-isp)#quit
huawei(config-aaa)#quit
huawei(config)#terminal user authentication-mode AAA isp
```

18.7 Configuration Example of the RADIUS AAA (802.1X access user)

The MA5800 is interconnected with the RADIUS server through the RADIUS protocol to implement authentication.

Service Requirements

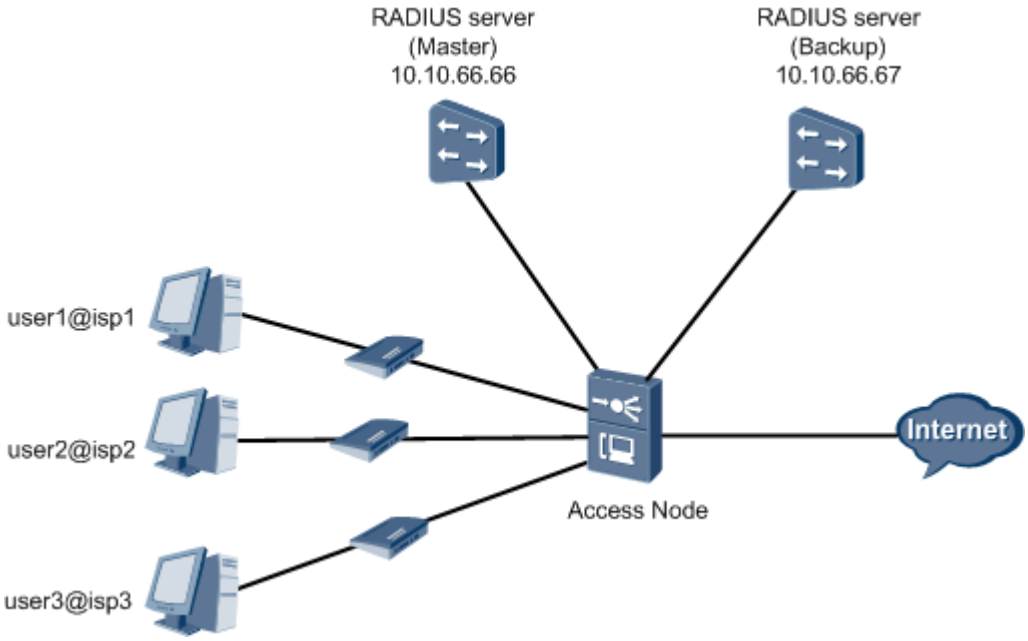
- The RADIUS server performs authentication for 802.1X access users of isp1.
- The user logs in to the server carrying the domain name.
- The RADIUS server with the IP address 10.10.66.66 functions as the primary server for authentication.
- The RADIUS server with the IP address 10.10.66.67 functions as the secondary server for authentication.
- Other parameters adopt the default settings.

Networking

Figure 18-5 shows an example network of the RADIUS authentication.

 Copyright © Huawei Technologies Co., Ltd.

Figure 18-5 Example network of the RADIUS Authentication and Accounting application.



Procedure

Step 1 Configure the authentication scheme.

Configure the authentication scheme named **newscheme** (users are authenticated through RADIUS).

```
huawei(config)#aaa
huawei(config-aaa)#authentication-scheme newscheme
huawei(config-aaa-authen-newscheme)#authentication-mode radius
huawei(config-aaa-authen-newscheme)#quit
huawei(config-aaa)#quit
```

Step 2 Configure the RADIUS protocol.

Create RADIUS server template named **hwtest** with the RADIUS server 10.10.66.66 as the primary authentication server, and the RADIUS server 10.10.66.67 as the secondary authentication server.

NOTE

Run the **(undo) radius-server user-name domain-included** command to configure the user name (not) to carry the domain name when transmitted to the RADIUS server.

- By default, the user name of the RADIUS server carries the domain name.
- After the **undo radius-server user-name domain-included** command is executed, the domain name is deleted from the user name when a client sends an authentication request to the RADIUS server.

```
huawei(config)#radius-server template hwtest
  Info: A new server-template is created.
  Warning: Please configure the shared-key. Configuring shared-key is mandatory to communicate with
RADIUS server.
huawei(config-radius-hwtest)#radius-server shared-key
huawei(config-radius-hwtest)#radius-server authentication 10.10.66.66 1812
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-radius-hwtest)#radius-server authentication 10.10.66.67 1812 secondary
huawei(config-radius-hwtest)#quit
```

Step 3 Configure the 802.1X authentication.

1. Enable the 802.1X global switch. Enable the 802.1X authentication for ports 1. The 802.1X needs to be triggered by DHCP. Therefore, the DHCP-trigger authentication must be enabled.

   ```
   huawei(config)#dot1x enable
   huawei(config)#dot1x service-port 1
   huawei(config)#dot1x dhcp-trigger enable
   huawei(config)#dhcp option82 enable
   ```

2. Configure an 802.1X parameters. In the local termination authentication, the 802.1X parameters should be configured to be in the EAP termination mode. The count of allowed handshake failure is 1 and the handshake interval is 20s.

   ```
   huawei(config)#dot1x keepalive retransmit 1 interval 20 service-port 1
   huawei(config)#dot1x eap-end service-port 1
   ```

Step 4 Create a domain.

NOTE

- A domain is a group of users of the same type.
- In the user name format userid@domain-name (for example, userdomain@domain.net), "userid" indicates the user name for authentication and "domain-name" followed by "@" indicates the domain name.
- The domain name for user login cannot exceed 20 characters.
- For a user with a domain name, the total length of the user name plus @ plus the domain name cannot exceed 31 characters.

Create a domain named isp1.

```
huawei(config)#aaa
huawei(config-aaa)#domain isp1
  Info: Create a new domain
```

Step 5 Use the authentication scheme.

You can use an authentication scheme in a domain only after the authentication scheme is created.

```
huawei(config-aaa-domain-isp1)#authentication-scheme newscheme
```

Step 6 Bind the RADIUS server template.

You can use a RADIUS server template in a domain only after the RADIUS server template is created.

```
huawei(config-aaa-domain-isp1)#radius-server hwtest
```

----End

Result

User 1 in isp1 can pass authentication only if both the user name and password are correct, and then can log in to the MA5800.

Configuration File

```
aaa
authentication-scheme newscheme
authentication-mode radius
```

 Copyright © Huawei Technologies Co., Ltd.

```
quit
quit

radius-server template hwtest
radius-server shared-key
radius-server authentication 10.10.66.66 1812
radius-server authentication 10.10.66.67 1812 secondary
quit
dot1x enable
dot1x service-port 1
dot1x dhcp-trigger enable
dhcp option82 enable
dot1x keepalive retransmit 1 interval 20 service-port 1
dot1x eap-end service-port 1

domain isp1
authentication-scheme newscheme
radius-server hwtest
```

18.8 Configuration Example of HWTACACS Authentication (Management User)

The MA5800 allows the management user of the device to log in to the system in the HWTACACS authentication mode.

Prerequisites

- The route from the MA5800 to the HWTACACS server must be configured.
- The management user information (user name@domain and password) must be configured on the HWTACACS server.

Service Requirements

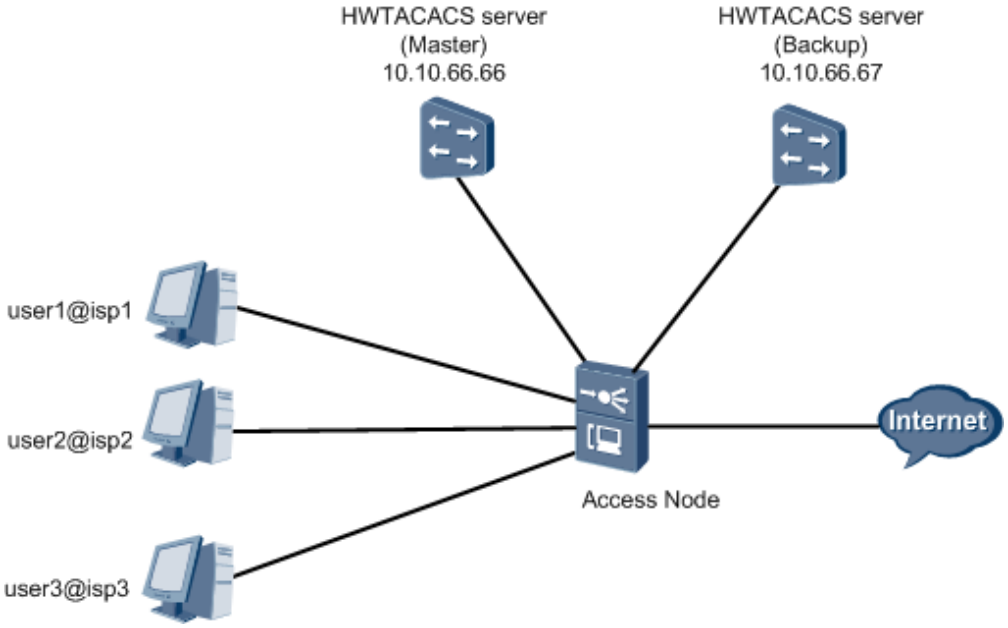
- The HWTACACS server performs authentication for management user of domain **isp1**.
- The user logs in to the server carrying the domain name.
- The HWTACACS server with the IP address 10.10.66.66 functions as the primary server for authentication.
- The HWTACACS server with the IP address 10.10.66.67 functions as the secondary server for authentication.
- Other parameters adopt the default settings.

Networking

Figure 18-6 shows an example network of HWTACACS authentication.

 Copyright © Huawei Technologies Co., Ltd.

Figure 18-6 Example network of HWTACACS authentication



Procedure

Step 1 Configure the authentication scheme.

Configure authentication scheme named **login-auth** (users are authenticated and authorized through HWTACACS).

```
huawei(config)#aaa
huawei(config-aaa)#authentication-scheme login-auth
huawei(config-aaa-authen-login-auth)#authentication-mode hwtacacs
huawei(config-aaa-authen-login-auth)#quit
huawei(config-aaa)#authorization-scheme login-auth
huawei(config-aaa-author-login-auth)#authorization-mode hwtacacs
huawei(config-aaa-author-login-auth)#quit
huawei(config-aaa)#quit
```

Step 2 Configure the HWTACACS protocol.

Create the HWTACACS server template named **hw-login** with HWTACACS server 10.10.66.66 as the primary authentication server, and HWTACACS server 10.10.66.67 as the secondary authentication server. Set the shared key.

```
huawei(config)#hwtacacs-server template hw-login
  Info: Create a new HWTACACS-server template.
  Warning: To improve the service security, please run the hwtacacs-server
shared-key command to configure a shared key.
huawei(config-hwtacacs-hw-login)#hwtacacs-server shared-key
huawei(config-hwtacacs-hw-login)#hwtacacs-server authentication 10.10.66.66 1812
huawei(config-hwtacacs-hw-login)#hwtacacs-server authentication 10.10.66.67 1812 secondary
huawei(config-hwtacacs-hw-login)#hwtacacs-server authorization 10.10.66.66 1812
huawei(config-hwtacacs-hw-login)#hwtacacs-server authorization 10.10.66.67 1812 secondary
huawei(config-hwtacacs-hw-login)#quit
```

Step 3 Create a domain named **isp1**.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

- A domain is a group of users of the same type.
- In the user name format userid@domain-name (for example, userdomain@domain.net), "userid" indicates the user name for authentication and "domain-name" followed by "@" indicates the domain name.
- The domain name for user login cannot exceed 20 characters.
- For a user with a domain name, the total length of the user name plus @ plus the domain name cannot exceed 31 characters.

```
huawei(config)#aaa
huawei(config-aaa)#domain isp1
  Info: Create a new domain.
```

Step 4 Use the authentication scheme **login-auth**.

You can use an authentication scheme in a domain only after the authentication scheme is created.

```
huawei(config-aaa-domain-isp1)#authentication-scheme login-auth
huawei(config-aaa-domain-isp1)#authorization-scheme login-auth
```

Step 5 Bind the HWTACACS server template **hw-login** to the user.

You can use an HWTACACS server template in a domain only after the HWTACACS server template is created.

```
huawei(config-aaa-domain-isp1)#hwtacacs-server hw-login
huawei(config-aaa-domain-isp1)#quit
huawei(config-aaa)#quit
```

----End

Result

- When the HWTACACS server is reachable, the management user can log in to the MA5800 through SSH. After entering the user name and password specified on the HWTACACS server, the management user can successfully log in to the MA5800.
- When the HWTACACS server is unreachable, a user cannot log in to the MA5800 in SSH mode by entering the user name and password specified on the HWTACACS server.

Configuration File

```
aaa
authentication-scheme login-auth
authentication-mode hwtacacs
quit
authorization-scheme login-auth
authorization-mode hwtacacs
quit
quit

hwtacacs-server template hw-login
hwtacacs-server shared-key
hwtacacs-server authentication 10.10.66.66
hwtacacs-server authentication 10.10.66.67 secondary
hwtacacs-server authorization 10.10.66.66
hwtacacs-server authorization 10.10.66.67 secondary
quit

aaa
```

 Copyright © Huawei Technologies Co., Ltd.

```
domain isp1
authentication-scheme login-auth
authorization-scheme login-auth
hwtacacs-server hw-login
quit
quit
```

 Copyright © Huawei Technologies Co., Ltd.

19 Device Management

This topic covers the overview, general specifications, availability, and sub-features of device management security.

19.1 ANCP
The Access Node Control Protocol (ANCP) is used by the broadband network gateway (BNG) to manage the line parameters (including QoS and user) of the access node (AN).

19.2 ANCP Proxy
In ANCP proxy networking, an ANCP proxy device is located between the broadband network gateway (BNG) and access network (AN) device. This ANCP proxy device is responsible for proxy forwarding of ANCP packets exchanged between the BNG and AN device.

19.3 LLDP
Link Layer Discovery Protocol (LLDP) is a standard link layer discovery mode defined in IEEE 802.1ab. LLDP deployed on multi-vendor devices that are running in different networks and managed by the network management system (NMS) enables adjacent network devices to exchange device information with each other. By accessing the device information, the NMS obtains detailed information, such as topology of the whole network and physical connections between devices, in real time. As such, LLDP helps users monitor network status and locate network faults.

19.4 NETCONF
Network Configuration Protocol (NETCONF) is a network configuration and management protocol based on Extensible Markup Language (XML). NETCONF enables the network management system (NMS) to quickly configure and manage remote devices.

19.5 Telemetry
Telemetry is a remote measuring technology that can remotely collect data from physical or virtual devices at a high speed. The Telemetry feature provides real-time and high-speed data collection and improves the device and network usage during data collection.

19.6 Inband Management VPN
Inband management VPN is a means by which carriers use the virtual private network (VPN) to manage and maintain devices and the management protocol on the device can use virtual routers for route forwarding.

 Copyright © Huawei Technologies Co., Ltd.

19.7 Remote Software Commissioning (Ethernet)
This section describes the implementation principles and configuration of remote software commissioning using Ethernet upstream transmission.

19.8 OLT ZTD

19.9 Power Saving

19.10 Version and Data Management
This topic provides an introduction to the version and data management feature, and describes the working principle of this sub feature.

19.11 Anti-theft

19.12 Environment Monitoring
This document describes the environment monitoring function of the MA5800 series products.

19.13 RTU

19.1 ANCP

The Access Node Control Protocol (ANCP) is used by the broadband network gateway (BNG) to manage the line parameters (including QoS and user) of the access node (AN).

19.1.1 ANCP Introduction

Definition

The Access Node Control Protocol (ANCP) is used by the broadband network gateway (BNG) to manage the line parameters (including QoS and user) of the access node (AN).

NOTE

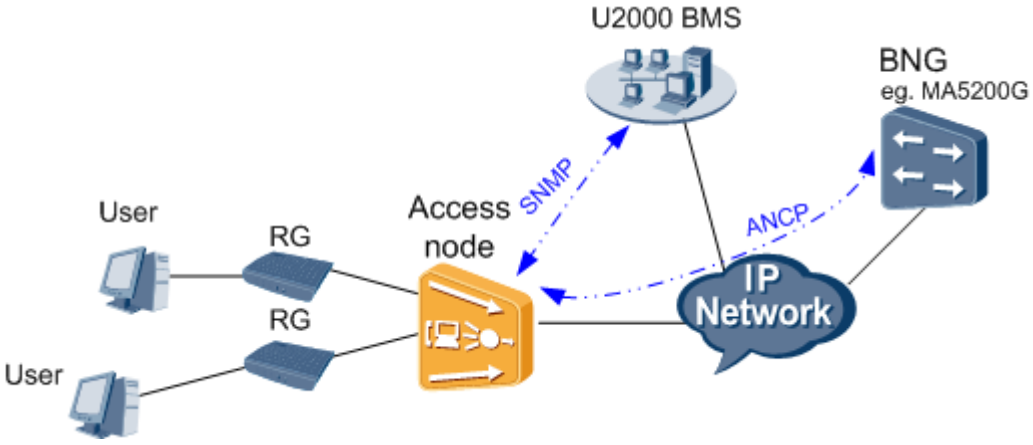
A BNG can be a BRAS or a router, such as the MA5200G.

Figure 19-1 displays the NEs relevant to ANCP.

- The user powers on, disables, or connects the RG to change the line status.
- The BNG and the AN exchange ANCP messages.
- The network administrator manages the AN through the N2000 BMS by using SNMP.

 Copyright © Huawei Technologies Co., Ltd.

Figure 19-1 ANCP network topology



Purpose

When ANCP is not used, if the BNG needs to manage the line parameters of an AN, the NMS is required. When the AN and the BNG use different NMSs, the line parameters are hard to be managed. Through ANCP, however, the BNG can directly manage such parameters without the NMS.

19.1.2 ANCP Principle

The ANCP supports the following functions:

- Line topology discovery
- L2C OAM
- Multiple partitions

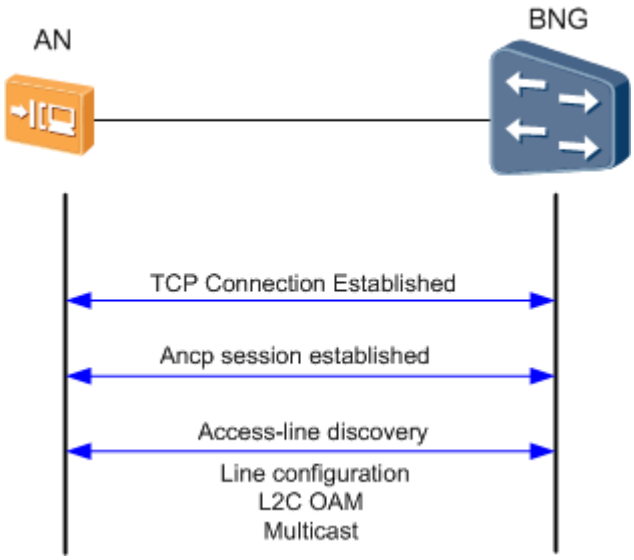
Before the above-listed ANCP functions are implemented, an ANCP session needs to be set up between the BNG and the AN.

Setting Up an ANCP Session

Figure 19-2 shows the process of setting up an ANCP session and negotiating the capabilities between the AN and the BNG.

 Copyright © Huawei Technologies Co., Ltd.

Figure 19-2 Setting up an ANCP session



The process of setting up an ANCP session is as follows:

1. On the AN side, pre-configure the ANCP session IP address and TCP port ID of the BNG and enable ANCP through the CLI. Then, the AN actively sends a request to the BNG for establishing a TCP connection (the BNG is the server and the AN is the client).
2. After the TCP connection is successfully established, adjacency is formed between the AN and the corresponding BNG. After the capabilities are negotiated, the ANCP session is successfully set up. If the local end finds that the remote end does not support a certain capability, the local end disables this capability and negotiates with the remote end again until both ends have negotiated the capabilities supported by both ends.
3. Configure capability parameters. Be default, the AN currently supports the capabilities of line topology discovery, line configuration, and L2C OAM. Multicast and unicast CAC can be added to the capabilities through configuration.

After the adjacency is set up, the ANCP protocol enters the maintaining stage. The AN handshakes with the BNG through the ACK message. The interval is the timeout time contained in the message exchanged during the adjacency setup process. If the AN does not receive the ACK message when the timeout time provided by the BNG expires for three times, the session between the AN and the BNG fails. The AN will then reset the adjacency and initiate a connection again.

Line Topology Discovery

The BNG records the actual parameter information about user ports through line topology discovery and thus implements QoS control.

After a line is activated and the port rate stabilizes, the ANCP module of the AN queries the parameters (such as upstream/downstream activation rate) of the line and sends the port up message and line parameter information to the BNG. After

 Copyright © Huawei Technologies Co., Ltd.

receiving the line information, the BNG saves the information to local and creates mapping to QoS control policies.

After the port is deactivated, the ANCP module sends the port down message to the BNG, as shown in **Figure 19-3**.

Figure 19-3 Line topology discovery

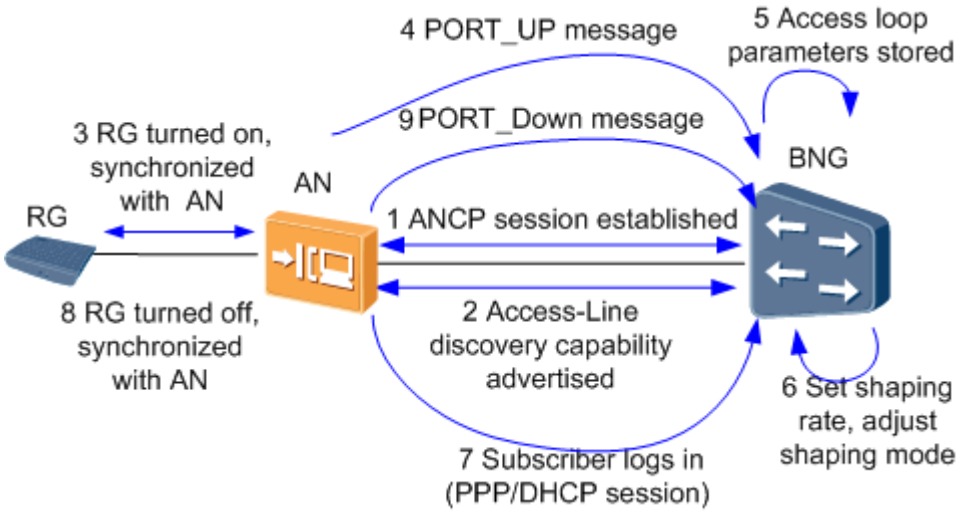


Table 19-1 lists the line parameters reported by the ANCP module.

Table 19-1 Reported line parameters

No.	Parameter
1	CID
2	RID
3	AGG-CID-ASCII
4	AGG-CID-BINARY
5	Actual-Net-Data-Rate-Upstream
6	Actual-Net-Data-Rate-Downstream
7	Line-State
8	Encapsulation
9	DSL Type

The port up or port down message is uniquely identified by the line ID. The format of the ANCP line ID is configurable. It is recommended to set the format of the ANCP line ID to be consistent with that of DHCP option 82 and PPPoE+ messages. At the same time, the format of the ANCP line ID must be the same as that on the BNG because the BNG creates mapping between user and line according to line ID and user name.

 Copyright © Huawei Technologies Co., Ltd.

The ANCP module can report the port up or port down message in two modes: based on port or based on service stream.

- In the port-based mode, if SPLABEL (configured by running the **raio-format** command) is not configured in the line ID, the following information is reported when the port is up or down:
  - Message in the default format (VPI=0, VCI=32) if the PVC or CVLAN is not specified
  - Specified value if the PVC or CVLAN specified exists
- If the port-based mode, if SPLABEL (configured by running the **raio-format** command) is configured in the line ID, the following situations occur when the port is up or down:
  - The topology information is not reported if the PVC or CVLAN specified does not exist.
  - The specified value is reported if the PVC or CVLAN specified exists.
- In the service-stream-based mode, when a port goes up or down, messages are reported for all service streams of the port. When the status of a service stream changes, a message is reported for this service stream.

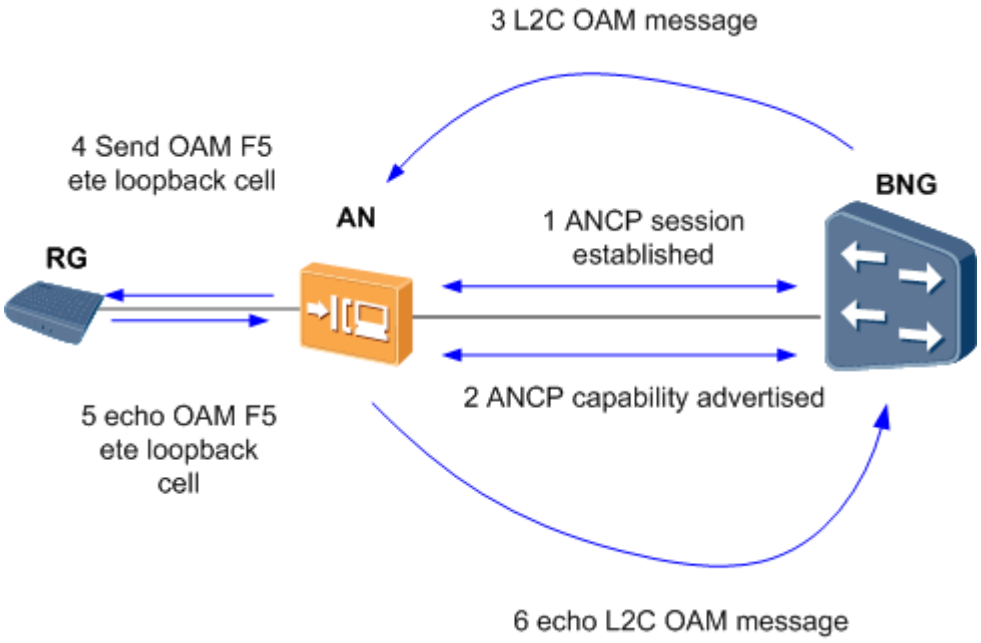
After the ANCP session fails and is re-established, the AN reports the stable line parameters and port status information, such as port UP or port DOWN, to the BNG.

L2C OAM

The ANCP L2C OAM function can implement the connectivity test between the BNG and the RG.

The implementation of the L2C OAM function mainly involves the RG, AN, and BNG, as shown in **Figure 19-4**.

Figure 19-4 L2C OAM



 Copyright © Huawei Technologies Co., Ltd.

L2C OAM implementation process:

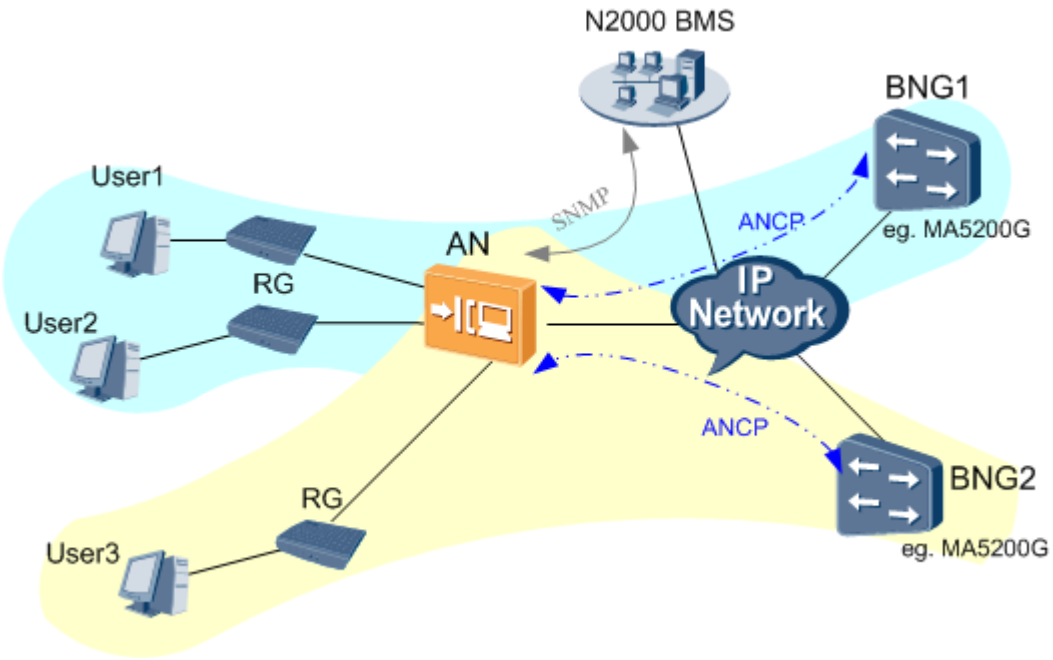
The AN sends response messages to the BNG according to the port status. If the port is up, the AN responds with a success message; if the port is down, the AN responds with a failure message.

Multiple Partitions

The ANCP multiple partitioning function enables different user ports to the grouped in different ANCP partitions. Different partitions are managed by different BNGs in order to support wholesale service. Generally, each BNG is managed by a respective ISP. One user port can belong to only one partition.

As shown in **Figure 19-5**, user 1 and user 2 are managed by BNG 1, and user 3 by BNG 2. In other words, user 1 and user 2 belong to the same partition, and user 3 belongs to a different partition.

Figure 19-5 Network topology of ANCP with multiple partitions



19.1.3 Configuring ANCP

Access Node Control Protocol (ANCP) is used to implement the functions such as topology discovery, line configuration, and L2C OAM on the user ports. The MA5800 establishes an ANCP session according to the GSMP communication IP address configured in the network access server (NAS).

Prerequisites

- The system must work in the normal state.

 Copyright © Huawei Technologies Co., Ltd.

- The system must be connected to the network access server in the normal state.

Context

- The MA5800 and the NAS use a TCP connection to carry an ANCP session. Therefore, before creating an ANCP session, you must create a TCP connection between the MA5800 and the NAS. The NAS functions as the server of the TCP connection, and the MA5800 functions as the client.
- After the TCP connection is created successfully between the MA5800 and the NAS, an ANCP session is created between the MA5800 and the NAS. After the ANCP session is created successfully, the MA5800 and the NAS need to use the ANCP ACK packets for heartbeat detection to maintain the ANCP session.
- The default values of the ANCP parameters are as follows:
  - GSMP address for an ANCP session: 0.0.0.0
  - ANCP session capability set: topology-discovery, line-config, and oam
  - ANCP packet sending priority: 6
  - GSMP TCP communication port number on the NAS side in an ANCP session: 6068
  - Interval for sending packets during the initial stage of an ANCP session: 10 (unit: 0.1s)
  - Interval for sending packets during the ANCP session stage: 250 (unit: 0.1s)

Procedure

Step 1 Run the **ancp partition enable** command to enable the ANCP partition function.

By default, the ANCP partition function is disabled.

Step 2 Run the **ancp port** command to enable the ANCP function of a port.

The ANCP function takes effect only when the ANCP function in the ANCP session mode and ANCP session function of a port are enabled.

Step 3 (Optional) Run the **ancp version** command to configure the ANCP version.

- The configured ANCP version must be the same as that on the NAS.
- By default, the ANCP version is draft-01.

Step 4 (Optional) Run the **raio-format ancp aggregation-circuit-id** command to customize the character string format for Access-Aggregation-Circuit-ID-ASCII TLV of the ANCP message.

NOTE

The system supports to customize the character string format for Access-Aggregation-Circuit-ID-ASCII TLV of the ANCP message only when the ANCP version is RFC6320.

Step 5 Run the **ancp session** command to create the ANCP session.

Step 6 (Optional) Run the **ancp partition** command to configure the ID of the partition associated with an ANCP session.

Step 7 Run the **ancp ip** command to configure the GSMP communication IP address for the ANCP session.

 Copyright © Huawei Technologies Co., Ltd.

- The IP address configured here must be the same as the GSMP communication IP address configured on the NAS, but it should to not be the same as the default IP address, multicast IP address, or broadcast IP address.
- When an ANCP session is enabled, the GSMP communication IP address of the NAS cannot be configured.

Step 8 (Optional) Run the **ancp capability** command to configure the capability set of the ANCP session.

- Supports topology discovery. When you select **topology-discovery** parameter, the MA5800 automatically reports the line parameters to the NAS.
- Supports line configuration. When you select **line-config** parameter, the MA5800 responds to the line configuration that is sent by the NAS.
- Supports the OAM. When you select **oam** parameter, the MA5800 responds to the line testing information that is sent by the NAS.
- Supports the preceding three types of capability.
- The default value is all, that is, the three capabilities (topology discovery, line configuration, and L2C OAM) are supported.

NOTE

When the ANCP version is RFC6320, the system supports only topology-discovery, and does not support line-config or oam.

Step 9 (Optional) Run the **ancp ancp-8021p** command to set the priority for sending ANCP packets.

- After an ANCP session is enabled, the priority for sending the ANCP packet of the ANCP session cannot be configured.
- You can set the priority according to the actual requirements and network conditions, the higher the priority, the higher the reliability.

Step 10 (Optional) Run the **ancp nas-tcp-port** command to set the GSMP TCP communication port number for the ANCP session on the NAS.

- By default, the GSMP TCP communication port number is 6068.
- The GSMP TCP communication port number on the MA5800 must be the same as that on the NAS.
- Run the **ancp port begin** command to set the start port ID of the ANCP session. Make sure that the start port ID of the ANCP session is the same as the start ID of the ports on the service board. The default value is 0.

Step 11 (Optional) Run the **ancp init-interval** command to set the interval for sending packets during the establishment of the ANCP session.

- By default, the general query interval is 1s.
- After an ANCP session is enabled, the priority for sending the ANCP packet of the ANCP session cannot be configured.

Step 12 (Optional) Run the **ancp keep-alive** command to set the interval for sending packets during the ACNP session so that the handshake messages can be sent to the peer end at the preset interval.

- After an ANCP session is enabled, the priority for sending the ANCP packet of the ANCP session cannot be configured.
- By default, the interval is 10s.

 Copyright © Huawei Technologies Co., Ltd.

Step 13 Run the **ancp enable** command to enable the ANCP session.

- Before an ANCP session is enabled, related parameters can be modified. After an ANCP session is enabled, related parameters cannot be modified.
- By default, the ANCP session is disabled.

Step 14 Run the **quit** command to quit the ANCP session mode.

Step 15 Run the **display ancp session** command to query the information about the ANCP session.

----End

Example

Consider configuring the ANCP topology discovery function of port 0/3/1 as an example. Configure the partition ID of the ANCP session to 1, ANCP version to rfc6320, the character string format for Access-Aggregation-Circuit-ID-ASCII TLV of the ANCP message is innervlanid and outervlanid, start port ID to 1, GSMP communication address of the ANCP session to 10.10.10.10, packet sending interval at the ANCP session creation phase to 2s, ANCP session capability set to topology-discovery, ANCP packet sending priority to 7, GSMP TCP communication port ID at the NSA side in the ANCP session to 6000, and packet sending interval at the ANCP session phase to 7s.

```
huawei(config)#ancp partition enable
huawei(config)#ancp port 0/3/1 partition 1
huawei(config)#ancp version rfc6320.
huawei(config)#raio-format ancp aggregation-circuit-id innervlanid.outervlanid
huawei(config)#ancp port begin 1
huawei(config)#ancp session 1
huawei(config-session-1)#ancp partition 1
huawei(config-session-1)#ancp ip 10.10.10.10
huawei(config-session-1)#ancp capability topology-discovery
huawei(config-session-1)#ancp ancp-8021p 7
huawei(config-session-1)#ancp nas-tcp-port 6000
huawei(config-session-1)#ancp init-interval 20
huawei(config-session-1)#ancp keep-alive 70
huawei(config-session-1)#ancp enable
huawei(config-session-1)#quit
huawei(config)#display ancp session 1
                              Session config status              : Enable
 Session running status             : Before syn phase
 Session diagnostic status          : -
 GSMP version                       : 3
 GSMP sub version                   : 1
 AN name                            : -
 NAS name                           : -
 NAS IP                             : 10.10.10.10
 Local IP                           : -
 AN instance                        : -
 NAS instance                       : -
 Config capabilities                : TopologyDiscovery
 Negotiate capabilities             : -
 NAS TCP port                       : 6000
 Startup time(0.01s)                : -
 Discontinuity time(0.01s)          : -
 Init interval(0.1s)                : 20
 Keepalive interval(0.1s)           : 70
 PartitionID                        : 1
  Line config roll default          : Disable
 OAM threshold(0.01)                : 100
 Topology report shaper interval(0.1s) : 10
 S-VLAN                             : -
```

 Copyright © Huawei Technologies Co., Ltd.

```
S-VLAN priority                  : 7
C-VLAN                           : -
C-VLAN priority                  : -
Session down send trap status    : Disable
Session up send trap status      : Disable
```

19.1.4 ANCP Reference Standards and Protocols

- draft-ooghe-ancp-framework-00
- draft-wadhwa-gsmp-l2control-configuration-00
- draft-wadhwa-gsmp-l2control-configuration-01
- draft-wadhwa-gsmp-l2control-configuration-02
- draft-decnodder-ancp-mib-an-01
- RFC3292
- RFC3293

19.2 ANCP Proxy

In ANCP proxy networking, an ANCP proxy device is located between the broadband network gateway (BNG) and access network (AN) device. This ANCP proxy device is responsible for proxy forwarding of ANCP packets exchanged between the BNG and AN device.

19.2.1 ANCP Proxy Introduction

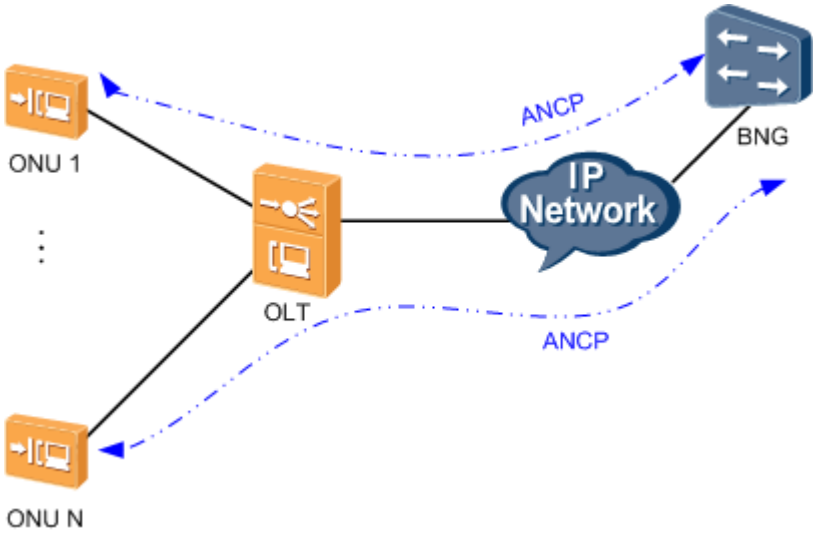
Definition

In ANCP proxy networking, an ANCP proxy device is located between the broadband network gateway (BNG) and access network (AN) device. This ANCP proxy device is responsible for proxy forwarding of ANCP packets exchanged between the BNG and AN device.

According to the standard Access Node Control Protocol (ANCP), a BNG needs to set up an ANCP session with every ONU. **Figure 19-6** shows a standard ANCP networking.

 Copyright © Huawei Technologies Co., Ltd.

Figure 19-6 ANCP networking



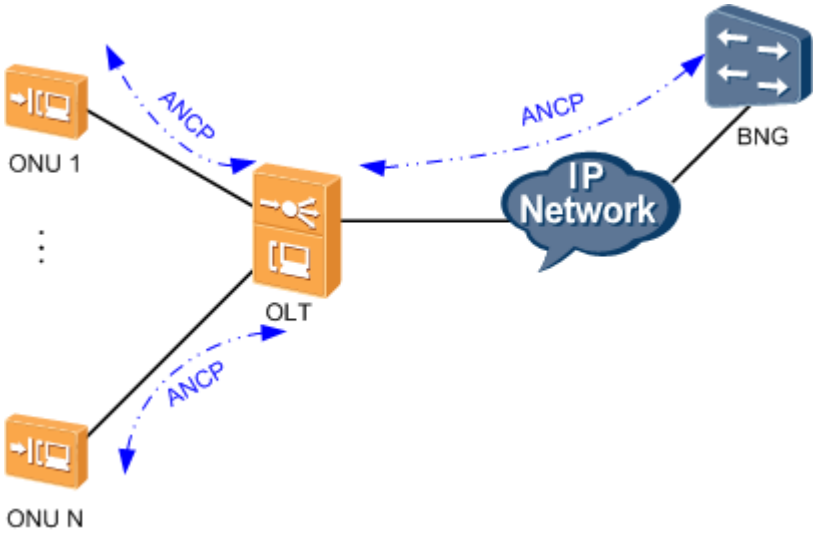
Purpose

ANCP sessions require period handshaking and TCP connections in ANCP sessions occupy memory resources. If lots of ONUs are deployed in an ANCP proxy network, ANCP sessions consume lots of BNG resources.

In an ANCP proxy network, the OLT functions as an ANCP proxy and ONU functions as an AN device. In this network, the BNG sets up an ANCP session with the ANCP proxy, and the ANCP proxy then sets up ANCP sessions with AN devices. In this case, the OLT converges the number of ANCP sessions, reducing the number of ANCP sessions on the BNG.

Figure 19-7 shows ANCP proxy networking.

Figure 19-7 ANCP proxy networking



 Copyright © Huawei Technologies Co., Ltd.

19.2.2 ANCP Proxy Principle

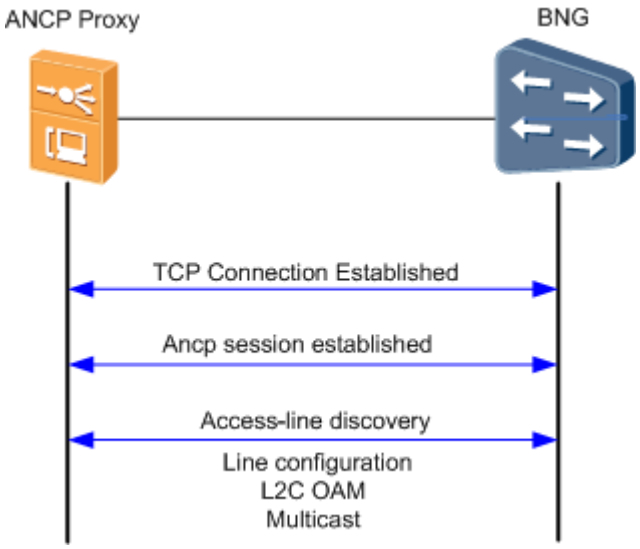
ANCP proxy is implemented by the following:

- Setting up a main ANCP session
- Setting up a sub ANCP session
- Setting up mapping between the main ANCP session and sub ANCP session
- Forwarding ANCP packets by proxy

Setting Up a Main ANCP Session

An ANCP session between the broadband network gateway (BNG) and ANCP proxy, that is, between OLTs, is called a main session. In the main ANCP session, ANCP proxy functions as the ANCP client. The main ANCP session negotiates capability sets between the BNG and ANCP proxy, as shown in **Figure 19-8**.

Figure 19-8 Setting up an ANCP session



The process of setting up a main ANCP session is as follows:

1. On ANCP proxy, pre-configure the ANCP session IP address and TCP port ID of the BNG and enable ANCP through the CLI. Then, the AN actively sends a request to the BNG for establishing a TCP connection (the BNG is the server and the ANCP proxy is the client).
2. After a TCP connection is set up, ANCP proxy establishes adjacency with the corresponding BNG. After capability set negotiation, an ANCP session is set up successfully. If the peer end does not support a capability, the local end accordingly disables the same capability. This process continues till both ends have the same capability set.
3. Configure capability set parameters. By default, ANCP proxy currently supports the capability set of line topology discovery, line configuration, and L2C OAM.

After the adjacency is set up, the ANCP protocol enters the maintaining stage. The AN handshakes with the BNG through the ACK message. The interval is the

 Copyright © Huawei Technologies Co., Ltd.

timeout time contained in the message exchanged during the adjacency setup process. If ANCP proxy does not receive the ACK message when the timeout time provided by the BNG expires for three times, the session between ANCP proxy and BNG fails. ANCP proxy will then reset the adjacency and initiate a connection again.

Setting Up a Sub ANCP Session

ANCP sessions between ANCP proxy and ONUs, that is, between the OLT and ANs, are sub ANCP sessions. In the main ANCP session, ANCP proxy functions as the ANCP client. When ANCP proxy and AN devices set up sub ANCP sessions, they negotiate about the capability set and ANCP partition ID on the AN device. The setup process of sub ANCP sessions is the same as that of the main ANCP session.

Only after the main ANCP session is set up can corresponding sub ANCP sessions be set up.

Setting Up Mapping Between the Main ANCP Session and Sub ANCP Session

ANCP proxy is used for establishing 1:N relationship between the main ANCP session and sub ANCP sessions so that the number of sessions can be reduced. Therefore, rules for mapping between the main ANCP session and sub ANCP sessions must be defined.

In mapping rules, the main ANCP session ID directly uses the existing ANCP session ID, and sub ANCP session IDs use AN IP addresses + partition IDs.

Forwarding ANCP Packets by Proxy

Proxy forwarding consists of the following 3 parts:

- Determination on upstream ANCP packet forwarding
- Determination on downstream ANCP packet forwarding
- Modification on ANCP packets

Determination on upstream ANCP packet forwarding

The main ANCP session has a 1:N relationship with sub ANCP sessions. After ANCP proxy receives upstream ANCP packets from a sub ANCP session, the unique main ANCP session can be determined. Then, ANCP packets are sent through this main session.

Determination on downstream ANCP packet forwarding

The main ANCP session has a 1:N relationship with sub ANCP sessions. After ANCP proxy receives downstream ANCP packets from the BNG, these packets can be forwarded according to other forwarding conditions.

- After receiving upstream ANCP packets, ANCP proxy extracts the AN ID (CID AN ID) from the packets and records the AN ID.
- After receiving downstream ANCP packets, ANCP proxy extracts the AN ID (CID AN ID) from the packets, and locates the corresponding sub ANCP session according to the AN ID and main session ID. Then, ANCP proxy sends these downstream ANCP packets.

Modification on ANCP packets

 Copyright © Huawei Technologies Co., Ltd.

The main ANCP session and sub ANCP sessions may have different partitions. Therefore, the Partition ID field needs to be modified for ANCP packet forwarding.

- For upstream ANCP packets, the Partition ID is modified to the partition ID.
- For downstream ANCP packets, the Partition ID is modified to the partition ID.

19.2.3 Configuring ANCP Proxy

Prerequisites

- The system must work in the normal state.
- The system must be connected to the network access server in the normal state.

Context

- ANCP proxy configuration requires main ANCP session configuration and sub ANCP session configuration. Before sub ANCP session configuration, the main ANCP session must be configured. For details about main ANCP session configuration, see **19.1.3 Configuring ANCP**. This topic describes how to configure a sub ANCP session.
- The MA5800 and the AN (ONU) use a TCP connection to carry an ANCP session. Therefore, before creating an ANCP session, you must create a TCP connection between the MA5800 and the AN. The MA5800 functions as the server of the TCP connection, and the AN functions as the client.
- After the TCP connection is created successfully between the MA5800 and the AN, an ANCP session is created between the MA5800 and the AN. After the ANCP session is created successfully, the MA5800 and the AN need to use the ANCP ACK packets for heartbeat detection to maintain the ANCP session.
- The default values of the ANCP parameters are as follows:
  - GSMP address for an ANCP sub session: 0.0.0.0
  - In an ANCP sub-session, the TCP communication port listened by the ANCP proxy is port 6068.
  - Interval for sending packets during the initial phase of an ANCP sub session: 10 (unit: 0.1s)
  - Interval for sending packets during the ANCP session phase: 250 (unit: 0.1s)

Procedure

Step 1 Run the **sysman server source ancp-proxy** command to bind a specified source port to the ANCP-Proxy service protocol.

NOTE

Step 2 Run the **ancp proxy** command to enable ANCP Proxy function.

Step 3 (Optional) Run the **ancp tcp-listen-port** command to configure the number of the TCP communication port listened by the ANCP proxy in the ANCP sub-session.

- The default value is 6068.

 Copyright © Huawei Technologies Co., Ltd.

Step 4 Run the **ancp sub-session** command to create the ANCP sub session.

Step 5 Run the **ancp partition** command to configure the ID of the partition associated with an ANCP sub session.

Step 6 Run the **ancp ip** command to configure the IP address for the ANCP sub session.

- The IP address configured here must be the same as the AN GSMP communication address, but it must not be the default IP address, multicast IP address, or broadcast IP address.
- When a session is enabled, the AN GSMP communication address cannot be configured.

Step 7 (Optional) Run the **ancp init-interval** command to set the interval for sending packets during the establishment of the ANCP sub session.

- The default value is 10s.
- After an ANCP sub session is enabled, the priority for sending the ANCP packet of the ANCP sub session cannot be configured.

Step 8 (Optional) Run the **ancp keep-alive** command to set the interval for sending packets during the ANCP sub session so that the handshake messages can be sent to the peer end at the preset interval.

- The default value is 250s.
- After an ANCP sub session is enabled, the priority for sending the ANCP packet of the ANCP sub session cannot be configured.

Step 9 Run the **ancp_session** command to configure the ANCP session associated with an ANCP sub-session.

Step 10 Run the **ancp enable** command to enable an ANCP sub session.

- Before an ANCP sub session is enabled, related parameters can be modified. After enabling, related parameters cannot be modified.
- The default value is disabled.

Step 11 Run the **quit** command to quit the ANCP sub-session mode.

Step 12 Run the **display ancp sub-session** command to query the information about the ANCP sub session.

----End

Example

Example:

- Bind the ANCP proxy service protocol to a specified source port. Set the port type to loopback and the port index to 1.
- Configure an ANCP sub-session. Set the number of the TCP communication port listened by the ANCP proxy function to 6000, ID of the ANCP sub-session to 1, ID of the associated partition to 1, GSMP communication address of the ANCP session to 10.10.10.10, interval for sending packets in the ANCP session establishment phase to 2s, interval for sending packets in the ANCP session phase to 7s, and ID of the associated main ANCP session to 1.

```
huawei#sysman server source ancp-proxy loopback 1
huawei(config)#ancp proxy enable
```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#ancp tcp-listen-port 6000
huawei(config)#ancp sub-session 1
huawei(config-sub-session-1)#ancp partition 1
huawei(config-sub-session-1)#ancp ip 10.10.10.10
huawei(config-sub-session-1)#ancp init-interval 20
huawei(config-sub-session-1)#ancp keep-alive 70
huawei(config-sub-session-1)#ancp session 1
huawei(config-sub-session-1)#ancp enable
huawei(config-sub-session-1)#quit
huawei(config)#display ancp sub-session 1
```

19.3 LLDP

Link Layer Discovery Protocol (LLDP) is a standard link layer discovery mode defined in IEEE 802.1ab. LLDP deployed on multi-vendor devices that are running in different networks and managed by the network management system (NMS) enables adjacent network devices to exchange device information with each other. By accessing the device information, the NMS obtains detailed information, such as topology of the whole network and physical connections between devices, in real time. As such, LLDP helps users monitor network status and locate network faults.

19.3.1 Introduction

This topic describes the background and application values of LLDP.

As large-scale networking requirements keep emerging, a network is involving an increasingly large number of device types, each device having its own complicated configurations. In addition, higher requirements are posed on the NMS capabilities. For example, the NMS is required to automatically learn the topological status of connected devices, and to detect configuration conflicts between devices.

Most NMSs use the automated discovery function to follow network topology changes, but in this way the NMSs can analyse the topology to as deep as the network layer only. As such, operators can learn only the basic events, such as device addition or deletion, on a network, but cannot identify the specific location of a device. For example, operators cannot identify the ports through which a device connects to other devices. Though some vendors provide proprietary protocols for discovering adjacency between devices, there are still not sufficient means for generating the topology of an entire network.

LLDP well addresses the issues above. As a Layer 2 discovery protocol defined in IEEE 802.1ab, LLDP provides a standard link layer discovery method. Using this method, information such as the capabilities, management address, device ID, and port ID of a local device can be encapsulated in LLDP frames and transmitted to adjacent nodes. After receiving such information, the adjacent nodes save the information in a standard management information base (MIB), which can be queried by the NMS for determining the connectivity of links.

LLDP can precisely discover the ports of network devices and the interconnection between devices. By obtaining and integrating the LLDP local device information and neighbor device information on each network element (NE), the NMS can generate a clear topology of the entire network, along with detailed information such as the physical connections between devices. Such information helps network

 Copyright © Huawei Technologies Co., Ltd.

operators monitor network status in real time and quickly locate network faults, effectively improving network security and stability.

19.3.2 LLDP Reference Standards and Protocols

This feature complies with the following standards and protocols:

- IEEE 802.1ab-2009
- IEEE 802.1ab-2005 (serves as the reference for the destination MAC address)

NOTE

IEEE 802.1ab-2009 provides a more complete definition of LLDP than IEEE 802.1ab-2005. Therefore, Huawei devices implement LLDP functions in compliance with IEEE 802.1ab-2009 by default. Specifically, LLDP IP addresses also comply with IEEE 802.1ab-2009.

19.3.3 Basic Concepts

This topic describes basic concepts related to the LLDP feature to help you better understand the working principles of LLDP.

LLDP MIB

MIB is a fundamental element for LLDP implementation. As defined in the LLDP protocol, each port on a device has four MIBs. Among them, the two most important MIBs are the LLDP local system MIB and the LLDP remote system MIB, which store the status information of the local device and the adjacent node, respectively. The status information includes device ID, port ID, system name, system description, port description, device capabilities, and network management address.

LLDP Agent

The LLDP protocol defines an LLDP agent for each port on a device for managing the LLDP operations regarding the device. The LLDP agent:

- Maintains the LLDP local system MIB information.
- Sends LLDP frames to the adjacent nodes, advertising the local system status information to the adjacent nodes.
- Recognizes and parses the LLDP frames sent from the adjacent nodes, and maintains the LLDP remote system MIB information.
- Sends LLDP alarms to the NMS when the LLDP local system MIB or LLDP remote system MIB information changes.

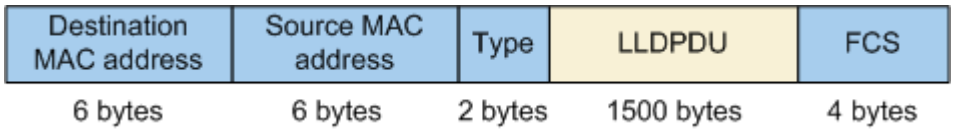
Compared with the IEEE 802.1ab-2009 specifications, Huawei's LLDP implementation supports one single port associated with one LLDP agent, but does not support one single port associated with multiple LLDP agents.

LLDP Frame

Ethernet frames in which LLDP data units (LLDPDUs) are encapsulated are called LLDP frames. The encapsulation has two formats: Ethernet II format and Subnetwork Access Protocol (SNAP) format. The MA5800 supports the Ethernet II encapsulation format. The following figure shows the structure of the LLDP frame encapsulated in this format.

 Copyright © Huawei Technologies Co., Ltd.

Figure 19-9 LLDP frame structure



The following table explains the fields of the LLDP frame.

Field	Description
Destination MAC address	Set to a fixed multicast MAC address 0x0180-C200-000E, which is specified by IEEE 802.1ab-2005. The address is specified by IEEE 802.1ab-2005. Therefore, for the purpose of maximal compatibility (that is, considering all devices as common devices), the MA5800 uses 01-80-C2-00-00-0E as the destination MAC address. For the requirements of different types of devices on the reception and transmission of the destination MAC address, see IEEE 802.1ab-2009.
Source MAC address	The device MAC address serves as the source MAC address.
Type	Indicates the frame type, which is a fixed value 0x88CC.
LLDPDU	Indicates the LLDP data unit. LLDP information exchange is implemented through LLDPDUs.
FCS	Indicates the frame check sequence.

LLDPDU

LLDPDU is a data unit that is encapsulated in an LLDP frame. A device encapsulates the local information in the Type-Length-Value (TLV) format. Several such TLVs constitute an LLDPDU and are transmitted in an LLDP frame. Users can formulate LLDPDUs using different combinations of TLVs based on their requirements. Using these TLVs, the device advertises its status and learns the status of its adjacent nodes.

Figure 19-10 LLDPDU structure

The LLDP protocol stipulates four mandatory TLVs for an LLDPDU. Each LLDPDU must start with the Chassis ID TLV, Port ID TLV, and Time to Live TLV, and end with the End of LLDPDU TLV. Other TLVs are optional.

There are two types of LLDPDUs:

- Normal LLDPDU: contains the local device information and the adjacent node information.

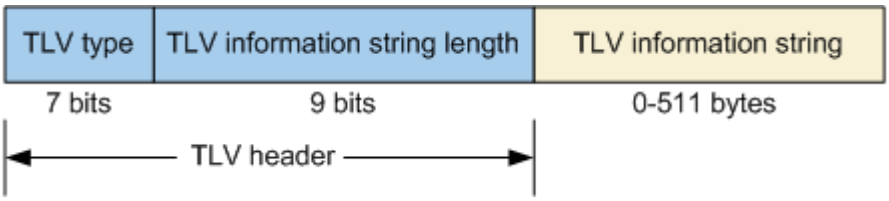
 Copyright © Huawei Technologies Co., Ltd.

- Shutdown LLDPDU: sent to disable the LLDP capability of a port. The shutdown LLDPDU is used to instruct the peer device to quickly clear the adjacent node information. The shutdown LLDPDU does not have optional TLVs, and the TTL TLV value is 0.

TLV

Indicating the type, length, and value of an object, TLV is the basic unit that forms an LLDPDU. Each TLV represents a type of device information, such as device ID, port ID, and management address, which correspond to fixed TLVs Chassis ID TLV, Port ID TLV, and Management Address TLV respectively.

Figure 19-11 TLV structure



LLDP supports two TLV formats: basic TLV and organizationally specific TLV. Organizationally specific TLVs include TLVs defined by 802.1 and 802.3, and may include more organizationally defined TLVs. The MA5800 supports only the basic TLV.

Table 19-2 TLV list

TLV Name	TLV Type	Description
End of LLDPDU TLV	0	Indicates the end of an LLDPDU.
Chassis ID TLV	1	Indicates the bridge MAC address of the device.
Port ID TLV	2	Indicates the name of the LLDPDU sending port. The value of this TLV uses the value of the ifName leaf of the IF-MIB.
Time To Live TLV	3	Indicates the time during which the local device information is valid on the adjacent node.
Port Description TLV	4	Indicates the port description. The value of this TLV uses the value of the ifDescr leaf in the iftable table of the IF-MIB.
System Name TLV	5	Indicates the device name, configurable using the **sysname** command.
System Description TLV	6	Indicates the device description, configurable using the **system sys-info description** command.

 Copyright © Huawei Technologies Co., Ltd.

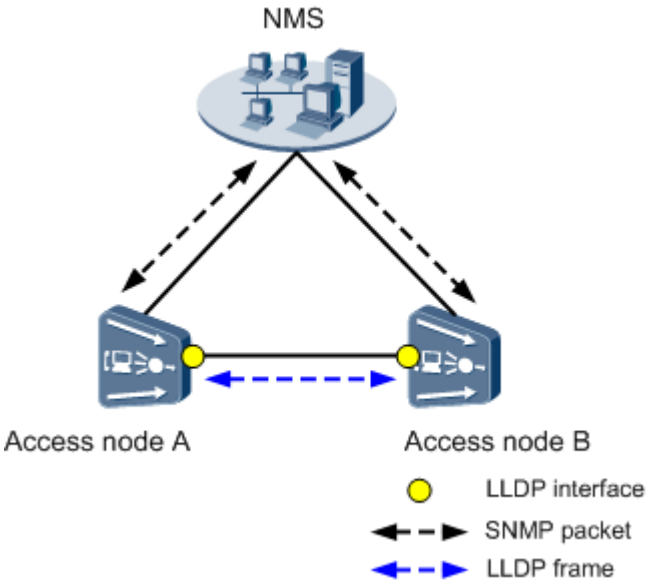
TLV Name	TLV Type	Description
System Capabilities TLV	7	Indicates the functions supported by the device and the functions that are enabled on the device.
Management Address TLV	8	Indicates the management address.
Reserved	9-126	Reserved for special application.
Organizationally Specific TLV	127	Each organization is represented by an organizationally unique identifier (OUI). For detailed description of this TLV and the OUI field, see the LLDP protocol.

19.3.4 Principles

Based on an LLDP basic network diagram, this topic describes how the NMS obtains device information and learns the network topology using the LLDP protocol.

LLDP Working Process

Figure 19-12 LLDP basic network diagram



As shown in the preceding figure, access node A and access node B support LLDP. The NMS collects device information in the following process:

1. Access node A encapsulates its local status information in an LLDP frame and sends the frame to its adjacent device, access node B.

 Copyright © Huawei Technologies Co., Ltd.

2. Access node B receives and parses the LLDP frame, and stores the information about access node A in its LLDP remote system MIB. The information can be accessed by the NMS when the NMS extracts the topology information.
3. Similarly, access node B encapsulates its local status information in an LLDP frame and sends the frame to access node A.
4. Access node A receives and parses the LLDP frame, and stores the information about access node B in its LLDP remote system MIB. The information can be accessed by the NMS when the NMS extracts the topology information.
5. The NMS exchanges SNMP messages with access node A and access node B, and extracts the local device information and adjacent device information from their MIBs. By integrating and analyzing the information, the NMS discovers the topology of the whole network.

In the working process described above:

- To generate the topology of the entire network on the NMS, all devices managed by the NMS must support LLDP. LLDP is disabled on the MA5800 by default.
- Each device can discover the information about only the device that is directly connected to this device. Therefore, to generate a whole-network topology, the NMS needs to collect the local device information and adjacent device information reported by all devices on the network.

LLDP Port Working Modes

An LLDP port supports the following working modes:

- Tx/Rx: The port transmits and receives LLDP frames. This is the default mode on the MA5800.
- Tx: The port transmits but does not receive LLDP frames.
- Rx: The port receives but does not transmit LLDP frames.
- Disable: The port does not transmit or receive LLDP frames.

When the LLDP working mode of a port changes, the port initializes the protocol state machine. If the LLDP working mode of a port changes frequently, the port has to constantly perform re-initialization. To avoid this situation, the MA5800 supports configuration of a port initialization delay. A port performs re-initialization only when the delay expires after the port working mode changes.

LLDP Frame Transmission Mechanism

An LLDP-enabled device periodically transmits LLDP frames to its adjacent nodes. To make sure that it is detected by other devices as quickly as possible, the MA5800 supports a fast transmission mechanism. The MA5800 immediately transmits an LLDP frame in any of the following three conditions:

- The MA5800 discovers a new neighbor. In other words, the MA5800 receives a new LLDP frame and finds that the transmitting device information carried in the frame is not stored on the MA5800 yet.
- The LLDP status of the MA5800 changes from disabled to enabled.
- The port status of the MA5800 changes from Down to Up.

 Copyright © Huawei Technologies Co., Ltd.

LLDP Frame Reception Mechanism

When receiving an LLDP frame, the device checks the validity of the frame and the TLV information carried in the frame. After the frame passes the validity check, the device stores the neighbor's information. According to the Time To Live (TLV) value carried in the LLDPDU, the device sets the aging time of the neighbor's information on this device. When receiving an LLDPDU whose Time To Live TLV value is 0, the device ages the information of the corresponding neighbor.

19.3.5 Network Application

The LLDP topology discovery function is usually used with the function of remote software commissioning (through GE upstream channel) to implement the plug-and-play (PnP) solution.

A general PnP solution is implemented in three steps:

1. After a device (connected to the upstream through GE) is powered on, the device automatically obtains its IP address and management VLAN parameters. Such information is used for setting up the management channel for the device to enable remote management.
2. Service configurations of the device are automatically issued to the device; automatic upgrade of the device is pre-deployed.
3. The network topology is discovered on the NMS.

The first step is implemented using the function of remote software commissioning (through GE upstream channel). With this function, operators need not visit the sites to commission software, thereby reducing the costs for deploying a large number of devices. Steps 2 and 3 require coordination of the NMS. The LLDP function works with the NMS to implement step 3.

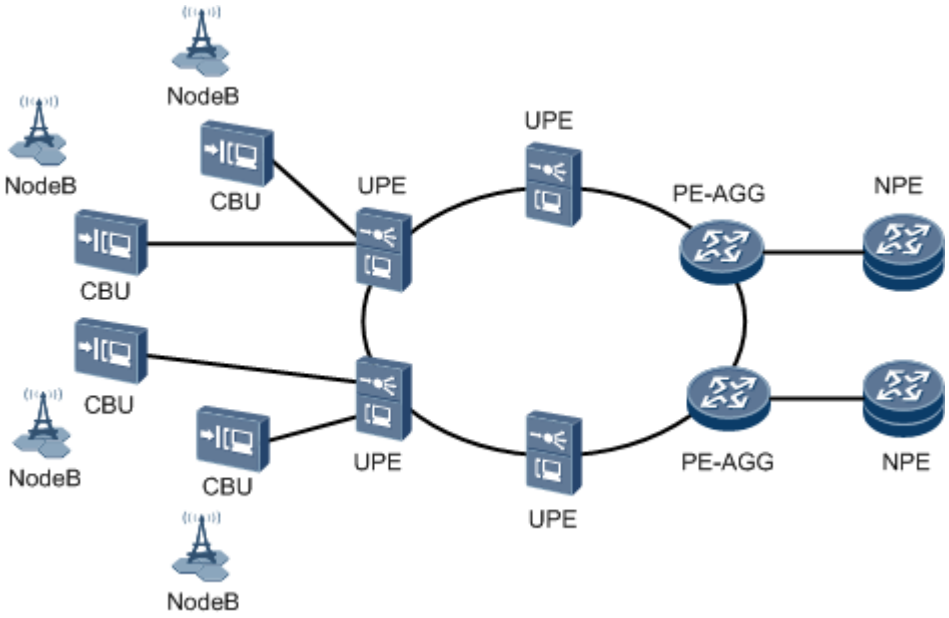
According to how the access device is connected to upstream devices, topology discovery of the access device includes the following scenarios.

Point-to-Point Connection

As shown in **Figure 19-13**, the access node MxU works as the CBU, and an OLT or a switch works as the UPE. The MxU connects to the upstream device through FE/GE. The IP address is obtained using DHCP, and the device location is identified by DHCP option 82. Specifically, DHCP relay or DHCP proxy is enabled on the upper-layer device, and the device reports the MxU location information to the NMS through DHCP option 82. The topology of this type of network can be discovered using LLDP.

 Copyright © Huawei Technologies Co., Ltd.

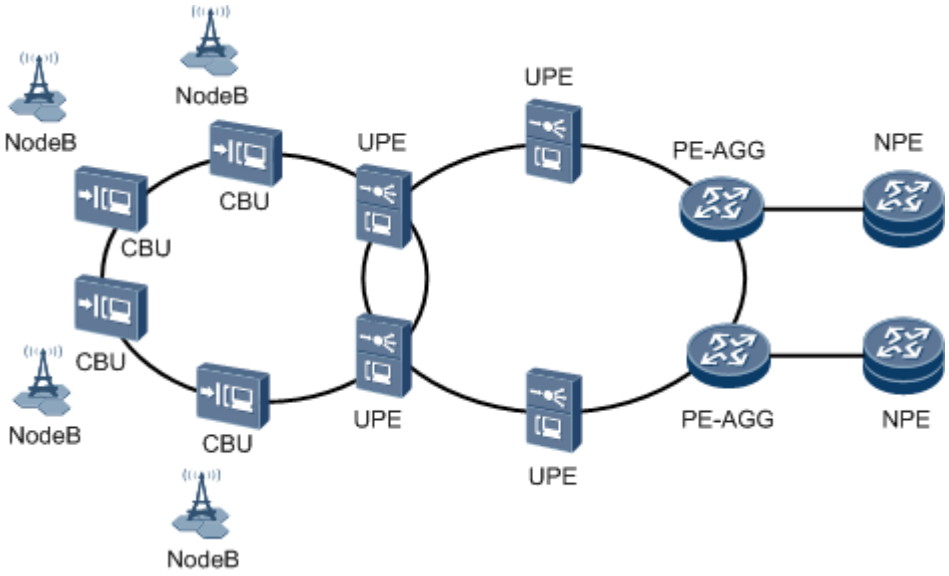
Figure 19-13 Network of point-to-point connection



Ring or Chain Network Topology

For a network topology shown in **Figure 19-14**, the device location cannot be identified using DHCP option 82. This is because DHCP option 82 does not apply to the scenario in which MxUs are cascaded. For such a network, LLDP can be used for topology discovery. Specifically, LLDP is used for discovering and creating the link topology information of adjacent devices. The NMS scans the link topology MIBs of all MxUs and their upper-layer devices, and builds the topology of the entire network. The location of an MxU can be identified through the connections between MxUs and their upper-layer devices. In this network scenario, the IP address is also obtained using DHCP. This network scenario applies to OLTs as well.

Figure 19-14 Ring network topology



 Copyright © Huawei Technologies Co., Ltd.

19.3.6 Configuring LLDP

If you need to manage devices of different vendors and networks, you can configure Link Layer Discovery Protocol (LLDP) on all the devices to obtain the detailed information, such as connections between network topologies and devices. This section describes how to configure LLDP on one device.

Prerequisites

All devices on a network support LLDP so that the network topology can be set up on the NMS.

Procedure

Step 1 Run the **lldp enable** command to enable the global LLDP function.

Step 2 Run the **lldp enable port** command to enable the port LLDP function.

LLDP must be enabled on all ports that are connected to devices. All LLDP-related configurations take effect only when both global LLDP and port LLDP are enabled.

Step 3 Run the **lldp management address** command to configure the LLDP management address.

Generally, the interface IP address for NMS management is set as the LLDP management address so the management interface IP addresses of devices can be queried using the LLDP topology.

Step 4 (Optional) Configure LLDP parameters.

NOTE

You are advised to configure the same LLDP parameters on all devices.

- Run the **lldp message-transmission interval** command to configure the interval between sending LLDP notification packets.
- Run the **lldp message-transmission hold-multiplier** command to configure the message hold multiplier. The message hold multiplier is used to calculate the TTL value of a sent packet. The system compares 65535 with hold-value x interval-value + 1, the smaller one is used as the TTL value.
- Run the **lldp restart-delay** command to configure the delay for switching the global or port LLDP status. The delay controls the minimum interval for switching the global or port LLDP status so that the system burden does not increase due to frequently changes of the status.

Step 5 (Optional) Configure LLDP trap function.

- Run the **lldp trap enable** command to enable the LLDP trap sending function.
- Run the **lldp trap interval** command to configure the interval for a port enabled with Link Layer Discovery Protocol (LLDP) to send traps. After the interval is configured, the device will not send a large number of traps to trouble the NMS or subscribers.

Step 6 Query the LLDP information.

- Run the **display lldp local** command to query the local LLDP information, including the system information, LLDP configuration, number of remote neighbors, and ports enabled with LLDP.

 Copyright © Huawei Technologies Co., Ltd.

- Run the **display lldp neighbor [port frameid/slotid/portid]** command to query the detailed neighbor information about the local device and port relationship between the neighbor device and local device.
- Run the **display lldp neighbor brief** command to query the brief neighbor information.
- Run the **display lldp statistics** command to query LLDP packet statistics.

Step 7 (Optional) Reset the LLDP information.

- Run the **reset lldp statistics** command to clear the packet statistics of all the ports or a specific port.
- Run the **clear lldp neighbor** command to clear the neighbor information about of all the ports or a specific port.

----End

Example

Set the management IP address to 10.10.10.2, set the interfaces connected to other devices to 0/9/0 and 0/9/1, use default values for other parameters. To configure LLDP and enable LLDP alarming, do as follows:

```
huawei(config)#lldp enable
huawei(config)#lldp enable port 0/9 0,1
huawei(config)#lldp management-address 10.10.10.2
huawei(config)#lldp trap enable
huawei(config)#display lldp local
{ <cr>|port<K>||<K> }:

  Command:
        display lldp local
  System information:
  ChassisId subtype               : macAddress
  ChassisId                       : 00e0-fc00-0000
  System name                     : MA5800
  System description              : Huawei Integrated Access Software
  System capabilities supported     : bridge
  System capabilities enabled       : bridge
  LLDP up time                    : 2021-08-21 20:39:03+08:00

  System configuration:
  LLDP status                     : enabled
  LLDP message TX interval          : 30s
  LLDP message TX hold multiplier   : 4
  LLDP restart delay              : 2s
  LLDP trap interval              : 30s
  LLDP trap enable                : enabled
  Management address              : 10.10.10.2

  Remote neighbors statistics:
  Remote neighbors last change time : 0 day(s),  0h: 0m: 0s
  Remote neighbors added            : 0
  Remote neighbors deleted          : 0
  Remote neighbors dropped          : 0
  Remote neighbors aged             : 0
  Total remote neighbors            : 0

  Port information:

  Port                          : 0/9/0
  PortId subtype                  : interfaceName
  PortId                        : ethernet0/9/0
  Port description                : Huawei-MA5800-ETHERNET
  LLDP enable status              : enabled TX-RX
```

 Copyright © Huawei Technologies Co., Ltd.

```
Total remote neighbors          : 0

Port                        : 0/9/1
PortId subtype                  : interfaceName
PortId                      : ethernet0/9/1
Port description                : Huawei-ETHERNET
LLDP enable status              : enabled TX-RX
Total remote neighbors          : 0

Total port number               : 2
```

19.4 NETCONF

Network Configuration Protocol (NETCONF) is a network configuration and management protocol based on Extensible Markup Language (XML). NETCONF enables the network management system (NMS) to quickly configure and manage remote devices.

19.4.1 Feature Updates

Version	Description
V100R019C1X and later versions	First version to support this feature.

19.4.2 Introduction

Network Configuration Protocol (NETCONF) is a key technology that simplifies complex network configuration.

Background

With the development of networks, the types and number of communication devices increase rapidly. Traditional Command Line Interface (CLI) or Simple Network Management Protocol (SNMP) configuration methods cannot adapt to the management of complex networks. Especially when devices from multiple vendors exist on the same network, the management methods of devices from different vendors are independent of each other, and the operations are complex. Therefore, automatic configuration is difficult to implement. To resolve this issue, the NETCONF protocol based on Extensible Markup Language (XML) and structured data model defined based on YANG are developed.

Definition

- **NETCONF** provides mechanisms to add, modify, delete, and query the configurations of network devices, and obtain the network device status. With NETCONF, network devices can provide standard application programming interfaces (APIs). Applications can directly use these APIs to send configurations to or obtain configurations from network devices. NETCONF can directly use the existing functional modules of devices. This reduces the development cost of NETCONF and allows users to use new features supported by the devices at any time.

 Copyright © Huawei Technologies Co., Ltd.

- **YANG** defines a structured data model. It is a machine-oriented model interface, and defines the data structure and constraints to describe data more flexibly and completely.

Value

NETCONF complies with YANG model constraints to implement communication between clients and servers. With standardized deployment, network administrators can use NETCONF to uniformly manage, configure, and monitor various network devices that support YANG, simplifying O&M and reducing O&M costs.

Compared with traditional configuration protocols, NETCONF is more convenient on man-machine interfaces and machine-to-machine interfaces. **Table 19-3** shows the comparison between SNMP and NETCONF.

Table 19-3 Comparison between NETCONF and SNMP

Comparison Item	SNMP	NETCONF
Configuration management	Does not provide the protection lock mechanism when multiple users perform operations on the same configuration.	Provides the protection lock mechanism to prevent conflicts caused by multi-user operations.
Query	Supports the query for one or more records in a table but requires multiple interactions.	Supports the direct query for the configuration data of the system, and supports filtering on data query.
Scalability	Poor expansibility	Good expansibility • SNMP models are defined by layer and each layer is independent of each other. Extensions to one layer have little effect on the other layers. • NETCONF is in the XML encoding format. This expands the management capability and system compatibility of NETCONF.

 Copyright © Huawei Technologies Co., Ltd.

Comparison Item	SNMP	NETCONF
Security	The SNMPv3 security protocol has no room for expansion.	NETCONF uses existing security protocols to ensure network security, and is not specific to any security protocols. This enables users to flexibly select a security protocol.

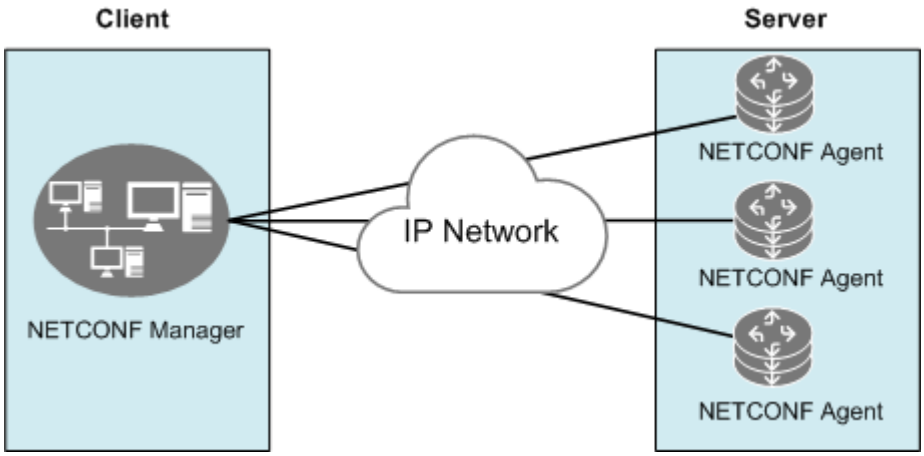
19.4.3 Principles

Network Configuration Protocol (NETCONF) uses a hierarchical structure. A lower layer provides services for the upper layer. The hierarchical structure enables each layer to focus only on a single aspect of NETCONF and reduces the dependencies between different layers.

NETCONF Network Architecture

NETCONF uses the Client/Server structure. **Figure 19-15** shows the basic network architecture.

Figure 19-15 NETCONF network architecture



The NETCONF network architecture comprises the following components:

- **NETCONF Manager**: serves as a client on the network and runs on the NMS/EMS. It works with the NETCONF Agent to implement device management. The network administrator uses the NETCONF Manager to send a **<rpc>** request in XML format to the NETCONF Agent.
- **NETCONF Agent**: serves as the server on the network. During device configuration, the NETCONF Manager sends a configuration management request to the NETCONF Agent. The NETCONF Agent parses the request and manages the device configuration and sends a response to the NETCONF manager in XML format.

 Copyright © Huawei Technologies Co., Ltd.

- **YANG**: NETCONF data modeling language, which is used to design configuration data and status data models, remote invocation models, and notification mechanisms for NETCONF operations.

NETCONF Framework Layers

NETCONF consists of four layers: content layer, operation layer, remote procedure call (RPC) layer, and application protocol layer. For details, see **Table 19-4**.

Table 19-4

Protocol Layer	Example	Description
Content layer	Configuration data	Describes the configuration data related to network management. The data varies depending on vendors' devices. The data modeling language is YANG.
Operation layer	<get-config>, <edit-config>, <notification>	Defines basic operations for applications in the RPC and calls these operations using XML encoding parameters.
RPC layer	<rpc>, <rpc-reply>	Provides a simple RPC request and response mechanism that does not depend on the transport protocol. • The client encapsulates the RPC request into an **<rpc>** element and sends the **<rpc>** element to the server. • The server encapsulates the request processing result in an **<rpc-reply>** element and sends the **<rpc-reply>** element to the client.

 Copyright © Huawei Technologies Co., Ltd.

Protocol Layer	Example	Description
Application protocol layer	BEEP, SSH, SSL	Provides the communication path between the NETCONF client and NETCONF server. Any transport protocol that meets the basic requirements can be used. The basic requirements for the transport protocol are as follows: • Connection-oriented. A persistent connection must be established between the NETCONF Manager and the NETCONF Agent. After the connection is established, reliable and serialized data transmission services must be provided. • Provides user authentication, data integrity, and security encryption for NETCONF. • The transport protocol provides a mechanism for distinguishing the session type (client or server) for NETCONF.

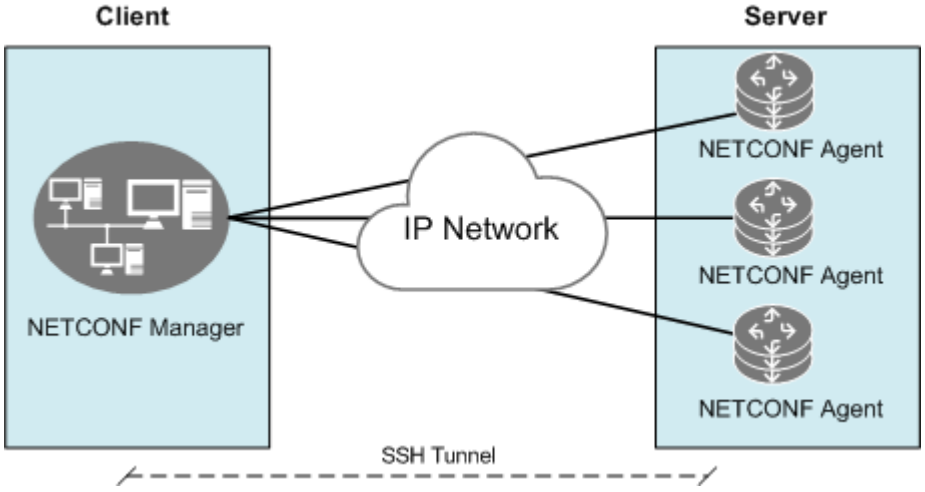
19.4.4 Networking Scenario

Network Configuration Protocol (NETCONF) ensures security and expansibility. When the NMS is used to manage network devices, you can use NETCONF to ensure communication between the NMS and the devices.

On a communications network, devices usually come from different vendors, and each vendor provides a unique device management method (for example, using different commands). In this scenario, configuring and managing these devices using traditional methods will be costly and ineffective. To resolve this issue, use NETCONF to remotely configure, manage, and monitor devices. The specified YANG file can be used to perform unified configuration. **Figure 19-16** shows the networking.

 Copyright © Huawei Technologies Co., Ltd.

Figure 19-16 NETCONF networking scenario



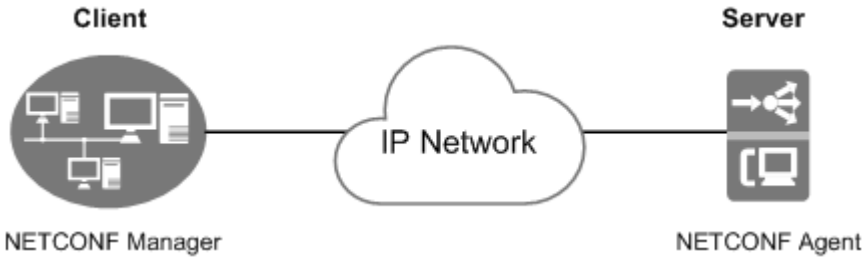
19.4.5 Configuring NETCONF

Configure Network Configuration Protocol (NETCONF) to implement remote communication between the network management system (NMS) and devices.

Networking Diagram

As shown in **Figure 19-17**, when NETCONF is configured, the NETCONF Manager functions as the client and the NMS is deployed. The NETCONF Agent functions as the SSH server and receives the connection from the NETCONF Manager that functions as the SSH client. In this way, the configuration file can be managed through NETCONF.

Figure 19-17 NETCONF Scenario



Procedure

Step 1 Run the **sysman server source** command to bind a specified source interface.

```
huawei(config)#sysman server source netconf vlanif 100
```

NOTE

In V100R020C00 and later versions and later versions, you must perform this step for the NETCONF function to take effect.

Step 2 Run the **sysman service netconf enable** command to enable NETCONF on the NETCONF agent and create a device user.

```
huawei(config)#sysman service netconf enable
 The netconf administrator user already exists, are you sure to create another?
```

 Copyright © Huawei Technologies Co., Ltd.

```
(y/n)[n]:y
 User Name(length<6,15>):user123
 User Password(length<12,128>):
 Confirm Password(length<12,128>):
 The operation is being processed. Please wait...
 The operation is successful
```

Step 3 Deploy the NMS on the NETCONF Manager.

Step 4 Log in to the NETCONF Agent from the NETCONF Manager using the NMS.

Step 5 Run the **display sysman server source netconf** command to check the configuration of the source interface.

```
huawei(config)#display sysman server source netconf
{ <cr>|ipv6<K>||<K> }:

 Command:
        display sysman server source netconf
 ------------------------------------------------------------------------------
 No.   InterfaceType   InterfaceNum   IPV4 Addrress    VPN Instance
 ------------------------------------------------------------------------------
 1     vlanif   -            -              -
 ------------------------------------------------------------------------------
```

NOTE

V100R020C00 and later versions support this command.

Step 6 Run the **display sysman service state** command to check the NETCONF configuration result.

```
 huawei(config)#
display sysman service state
{ <cr>||<K> }:

 Command:
        display sysman service state
 ------------------------------------------------------------------------------
 Network service                          Port          State
 ------------------------------------------------------------------------------
 telnet                                   23            disable
 trace                                    1026          disable
 ssh                                      22            enable
 snmp                                     161           enable
 ftp-client                               ----          ----
 sftp-client                              ----          ----
 ntp                                      123           enable
 radius                                   ----          enable
 dhcp-relay                               67            disable
 dhcpv6-relay                             547           disable
 ntp6                                     123           disable
 ipdr                                     4737          enable
 twamp                                    862           enable
 netconf                                  830           enable
 telnetv6                                 23            disable
 sshv6                                    22            disable
 snmpv6                                   161           disable
 web-proxy                                8024          disable
 portal                                   2000          disable
 capwap                                   5246          disable
 ------------------------------------------------------------------------------
```

----End

 Copyright © Huawei Technologies Co., Ltd.

19.4.6 NETCONF Reference Standards and Protocols

Table 19-5 Main reference standards of NETCONF

Standard No.	Description
RFC 4741	NETCONF Configuration Protocol
RFC 4742	NETCONF Configuration Protocol over Secure Shell (SSH)
RFC 6241	Network Configuration Protocol (NETCONF)
RFC 6242	Using the NETCONF Protocol over Secure Shell (SSH)
RFC 5277	NETCONF Event Notifications
RFC 6536	Network Configuration Protocol (NETCONF) Access Control Model

19.5 Telemetry

Telemetry is a remote measuring technology that can remotely collect data from physical or virtual devices at a high speed. The Telemetry feature provides real-time and high-speed data collection and improves the device and network usage during data collection.

NOTICE

According to industry practices, Telemetry data collection may involve obtaining personal data and personal communication content (the product does not store, parse, or process such data) for the purpose of ensuring network operation and services. Huawei alone is unable to collect or save the personal data of and user communication content. It is suggested that you activate the interception-related functions based on the applicable laws and regulations in terms of purpose and scope of usage. You are obligated to take considerable measures to ensure that the personal data of users and the content of users' communications are fully protected when the personal data and the content are being used and saved.

19.5.1 Telemetry Feature Updates

Product Version	Change Description
V100R020C00	First version to support this feature.

Copyright © Huawei Technologies Co., Ltd.

19.5.2 Introduction to Telemetry

Context

As the network expands, more and more services are carried. The intelligent network O&M system poses higher requirements on the number, performance, and precision of network monitoring devices.

Traditional network monitoring (SNMP and syslog) uses the pull mode to periodically poll each device on the network to obtain device data, resulting in high pressure on devices and low query efficiency when massive data needs to be queried. In addition, different vendors have different data formats and unstructured data, making it difficult to adapt analysis tools. Therefore, a new network monitoring mode is urgently needed.

Definition

Telemetry, as a mainstream performance collection mechanism for the future, is a network monitoring mode in push mode. That is, sampled devices on the network proactively push the interface traffic data to the NMS periodically. The push mode of Telemetry can achieve second-level or subsecond-level sampling precision.

Benefits

The Telemetry network monitoring mode enables the intelligent network O&M system to monitor more devices with higher precision. In addition, the monitoring process has little impact on the functions and performance of devices.

Compared with traditional SNMP query, Telemetry has higher collection efficiency. **Table 19-6** describes the advantages of Telemetry sampling compared with SNMP query.

Table 19-6 Comparison between the SNMP query and telemetry sampling

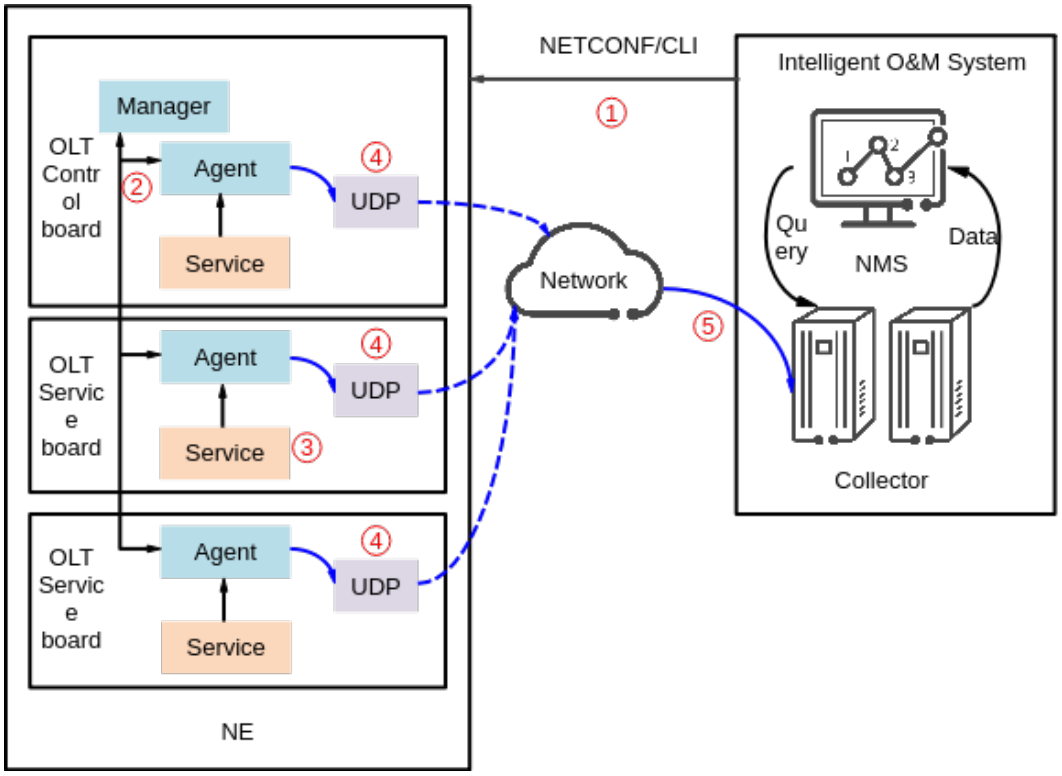
Telemetry		SNMP	
Push collection	NEs continuously push collected data after one subscription.	Pull collection	The NMS needs to send a query request each time data is collected.
Distributed collection	Interface boards send data directly.	Centralized collection	The collected data is aggregated to the control board.
Hardware collection	Hardware-based data collection allows data to be collected within milliseconds.	Software collection	Software-based data collection has a long collection time.
Google Protocol Buffer (GPB)	Improves storage and transmission efficiency.	Unstructured data	Low information processing efficiency

 Copyright © Huawei Technologies Co., Ltd.

19.5.3 Telemetry Data Collection and Reporting Process

In the application scenario where access network devices support the Telemetry interface-based collection function, the Telemetry collection system can be divided into the intelligent O&M system and OLT. The intelligent O&M system makes collection configurations, receives data collected by the OLT through Telemetry, and processes and analyzes the data. The OLT reports data collected through Telemetry based on the collection configurations.

Figure 19-18 Data collection process



1. The intelligent O&M system uses NETCONF/CLI static configurations to deliver the collection model to NEs that support Telemetry. Collection models include:
   - Data source type
   - Sampling frequency
   - Upstream transmission channel
   - Data format
2. The OLT control board delivers the collection and subscription information to specific service boards.
3. Service boards sample data (such as interface traffic) or monitor alarms and events and store the data.
4. Upload the collected data through UDP.
5. The collector collects sampled data.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

- The collector is a big data analysis system that is capable of receiving, storing, and processing sampled data reported by devices on the network.
- The intelligent O&M system processes services based on the analysis results from the big data analysis system, for example, checking the network status and optimizing the network promptly. After an optimization instruction takes effect, newly sampled data is sent to the collector. The intelligent O&M system can analyze whether the optimization effect meets expectations.

19.5.4 Telemetry Data Model

The collection controller consists of the control module and collection module. The control module must support the NETCONF protocol and configurations based on the openconfig-telemetry.yang and ×××-openconfig-telemetry-ext.yang data models. The OLT can function as a UDP agent and encapsulate and report data based on ×××-telemetry.proto and the data collection model.

Table 19-7 NETCONF configuration protocol data model

Layer	Description
Content layer	Includes the collection configuration data. The openconfig-telemetry.yang and ×××-openconfig-telemetry-ext.yang YANG models are used to model the collection configuration data, and the sensor-path contained in the models is used to specify collection items. The configuration must comply with IETF RFC 7950.
Operation layer	Supports NETCONF. NETCONF implementation must comply with IETF RFC 6241.
Messages layer	Supports NETCONF. NETCONF implementation must comply with IETF RFC 6241.
Communication protocol layer	Supports NETCONF. NETCONF implementation must comply with IETF RFC 6241.

The collected packets reported by the OLT to the collection module must comply with the UDP protocol. The following table lists the protocol layers and requirements.

 Copyright © Huawei Technologies Co., Ltd.

Table 19-8 Protocol data model for packet collection in UDP reporting mode

Layer		Description
Data model layer **NOTE** The data model layer is divided into two layers, which are defined in different .proto files.	Service data layer	Complies with the definition of the sensor_path field in the ×××-telemetry.proto file. The value of this field is obtained from different service data .proto files. 0 is meaningful data and should be reported. All Fs indicate data that cannot be reported due to device faults.
	Telemetry layer	The value must comply with the definition in the ×××-telemetry.proto file.
Message header layer		(Optional) Fragmentation and coding format indication can be implemented through the message header layer. In the message header layer, bit 4 is the S bit, and bits 5 to 8 are ET bits. When S is set to 1 and ET is set to 0, the encoding format is GPB. For details, see IETF draft-ietf-netconf-udp-pub-channel-05.
UDP layer		Supports the UDP protocol and complies with IETF RFC768 when implementing UDP.

The collection items supported by the OLT, collection performance requirements on a single collection item, and collection data model are defined in the service proto file.

NOTE

The sensor path format in the collection configuration must meet the following requirements:

- Data model file name at the service data layer: message name/.../message name
- message name/message name/message name indicates the node path in the data model. Set this parameter based on the site requirements.
- When a single collection object needs to be collected, the filter needs to be configured in the huaweiopencon-telemetry-ext.yang file.

  Example: The message QueueKpiRecords of the an-bb-quene-kpi.proto model is used to collect queue traffic, and the sensor path is an-bb-quene-kpi:QueueKpiRecords.

19.5.5 Key Telemetry Technologies

19.5.5.1 UDP Transport Protocol

UDP is a simple datagram-oriented transport layer protocol that does not provide reliability, flow control, or error recovery functions for IP. UDP applies to networks that have low reliability requirements but require economical transmission.

 Copyright © Huawei Technologies Co., Ltd.

Telemetry uses UDP to report the data encapsulated in an encoding format to the collector.

NOTE

For the definition of the UPC message header, see draft-ietf-netconf-udp-pub-channel-05.

Figure 2 shows the layers of the UDP protocol stack.

Figure 2 UDP protocol stack layers

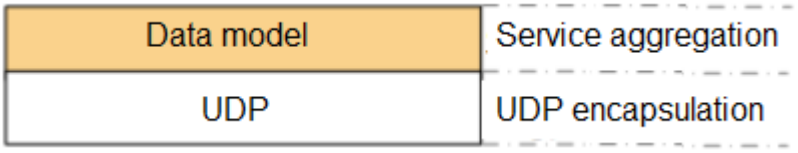


Table 19-9 describes the layers.

Table 19-9 Table 2 Description of each layer

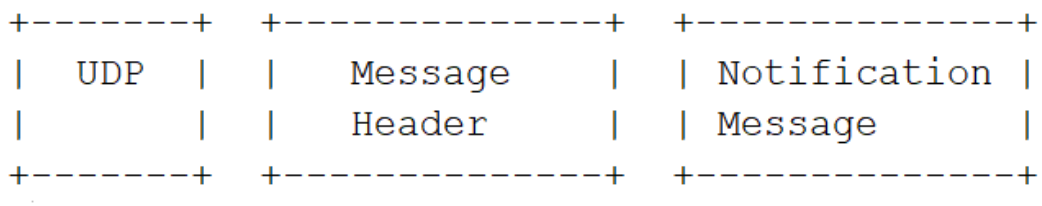
Layer	Description	Remark
UDP layer	Communication protocol based on the UDP connection.	-
Message header layer	For the structure definition of the message header layer, see **UPC Message Header**.	The data model layer is divided into three layers. Except the message header layer, the other two layers are defined in different .proto files.To establish a UDP connection between the collector and a network device, use the .proto files for interconnection. The .proto files on the two ends must be the same to facilitate decoding.
Telemetry layer	When the encoding format of the data to be sent is GPB, it is defined in the **huawei-telemetry.proto** file.	
Service data layer	When the encoding format of the data to be sent is GPB, the value of the **encoding** field in the **huawei-telemetry.proto** file is Encoding_GPB (0), the **data_gpb** field carries sampled data in GPB format, and the **data_str** field is empty. **NOTE** The **data_gpb** field needs to be decoded using the .proto file of the corresponding service, which is determined by the **sensor_path** field in the **huawei-telemetry.proto** file. For example, if the value of the **sensor_path** field is **huawei-ifm:ifm/interfaces/interface**, the .proto file is **huawei-ifm.proto**.	

UPC Message Header

Figure 19-19 describes the typical UPC message structure.

 Copyright © Huawei Technologies Co., Ltd.

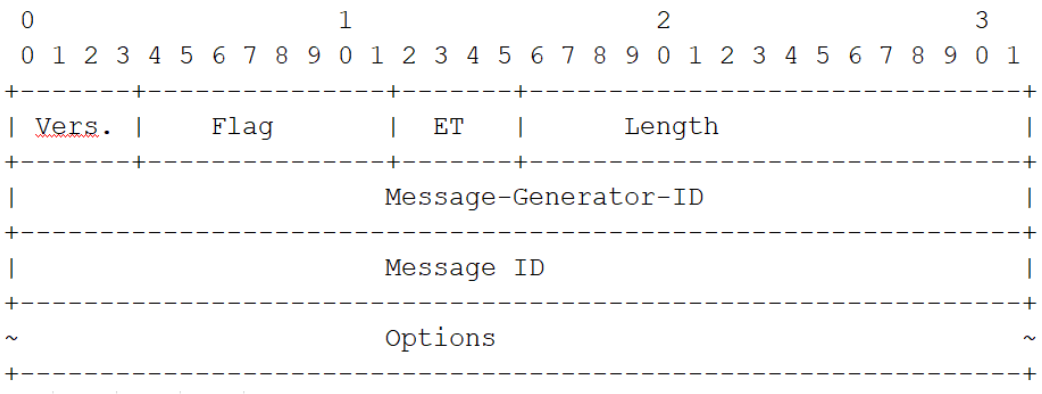
Figure 19-19 UDP release message description



- A message header packet includes information that can facilitate message transmission before deserializing a notification message.
- A notification message is the encoded content transmitted through a release channel. Common encoding methods include GPB, CBOR [RFC7049], JSON, and XML.

Figure 19-20 describes the data format of the UPC message header.

Figure 19-20 Data format of the UPC message header



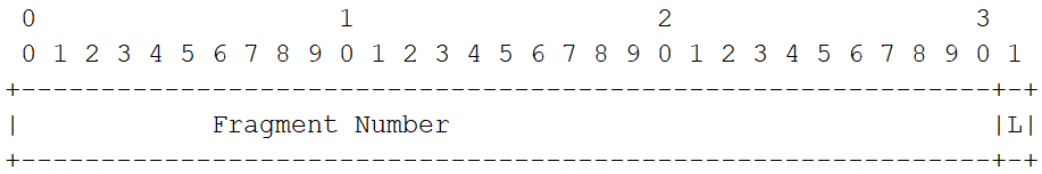
- A UPC message header occupies only 12 bytes, excluding the Options field, if the message is not fragmented.
- Vers.: indicates the encoding version of the protocol data unit (PDU). The initial value is 0.
- Flag: a bitmap that indicates the functions and additional options of a data packet. Each bit is associated with a function and an option data. When a bit is set to 1, the associated function is enabled and option data is added. The order in which options exist follows the bit order of the bitmap. In this document, the flags are specified as follows.
  - bit 0, the reliability flag;
  - bit 1, the fragmentation flag;
  - Other bits are reserved.
- ET: a 4-bit identifier that indicates the encoding type used for the notification message. 16 encoding types can be expressed.
  - 0: GPB
  - 1: CBOR
  - 2: JSON

 Copyright © Huawei Technologies Co., Ltd.

  - 3: XML
  - Others are reserved
- Length: indicates the total length of a message, in octets, including the message header.
- Message-Generator-ID: 32-bit identifier of the process that creates the notification message. This allows ambiguity to be removed from the source of information, such as identifying different line cards that send notification messages. The source IP address of the UDP datagram should not be interpreted as the identifier of the host initiating the UPC message. The entity that sends a UPC message may be only a relay.
- Message ID: It is generated by the message generator consecutively. Different users share the same notification ID sequence.
- Options: variable-length field. Details about options are described in the following sections.

The following figure shows the format of the option part. When the fragment flag is set to 1, the fragment option is available in the message header.

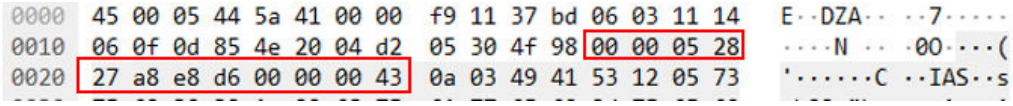
Figure 19-21 Fragment option format



- Fragment Number: indicates the sequence number of the current fragment.
- L: a flag used to indicate whether the current fragment is the last fragment. When this parameter is set to 0, the current fragment is not the last fragment. Therefore, more fragments are expected. If this parameter is set to 1, the current fragment is the last fragment.

UDP Message Example

Figure 19-22 UPC message header example



- The content in the red box is the content of the UPC message header.
- First uint32:
  - vers = 0
  - flag = 0: no option, that is, no fragment.
  - ET = 0: The reported content is in GPB coding format.
  - length = 0x0528 = 1320: same as the UDP payload length parsed by the Wireshark. Figure 1-6
- Second uint32:

 Copyright © Huawei Technologies Co., Ltd.

  - Message-Generator-ID = 0x27a8e8d6: reported process ID. The MA5800 reports the 1970s timestamp.
- Third uint32:
  - Message ID = 0x43: message sequence number. The value is incremented by 1 each time a packet is sent.
- All bytes following 0x0a starting from the 0x28 address are GPB codes.

19.5.5.2 GPB Encoding Format

Telemetry uses the GPB encoding format (the file name extension of a file encoded in GPB format is .proto) and provides a mechanism for serializing structured data flexibly, efficiently, and automatically. The GPB format is binary encoding and has good performance and high efficiency.

- GPB encoding depends on the information about the data structure. The message data structure used by the GPB is defined in the proto file. The message definition in the proto file can be converted from a yang file or manually compiled. Each field in the message data structure definition has a unique index. Once the proto is released, the index of each field in the message data structure should not be changed.
- GPB encoding uses a series of key-value pairs to represent fields in a message data structure.
- The key is obtained by combining the field index and data type of each field in each message, for example, key = (index<<3) | type.
- In GPB encoding, seven bits in a byte indicate valid data, and the most significant bit indicates whether the current byte is the last byte.
- Because the type in the key occupies 3 bits, if the index value is large, more bytes are required to represent the key.
- For example, a field index ranging from 1 to 15 requires only one byte for GPB encoding, including the field index and type (you can find more information in Protocol Buffer Encoding). A field index ranging from 16 to 2047 occupies two bytes, and a larger index requires more bytes. Because of this GPB encoding representation, indexes 1 to 15 should be reserved for message elements that appear very frequently. GPB allows indexes ranging from 1 to 2^{29} – 1 (536,870,911). Indexes 19000 to 19999 are reserved for GPB and cannot be used.

Types Supported by GPB Encoding

Table 19-10 Types supported by GPB encoding

Type	Meaning	Used For
0	Varint	int32, int64, uint32, uint64, sint32, sint64, bool, enum
1	64-bit	fixed64, sfixed64, double
2	Length-delimited	string, bytes, embedded messages, packed repeated fields

 Copyright © Huawei Technologies Co., Ltd.

Type	Meaning	Used For
3	Start group	groups(deprecated)
4	End group	groups(deprecated)
5	32-bit	fixed32, sfixed32, float

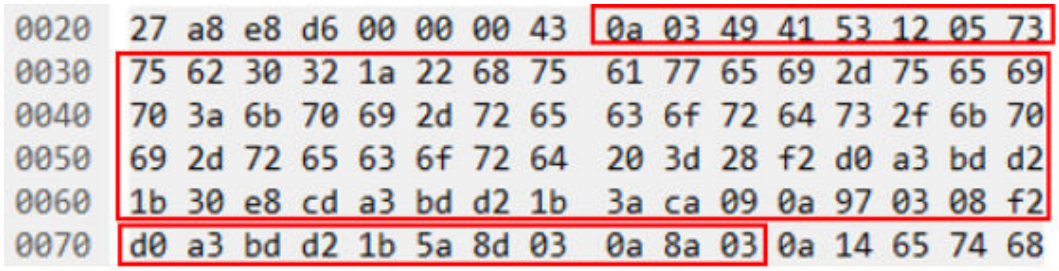
- The preceding types 3 and 4 are obsolete.
- The remaining types 0, 1, and 5 are all numeric types in the format of key-value.
- The remaining type 2 is a variable-length type, and is used to indicate string, bytes, message nesting, and the like in the format of key-length-value. If the value is a string, \0 at the end is not included.

Base 128 Varints encoding (compression algorithm) is used for keys, numeric values, or lengths: A compact method for representing numbers based on the assumption that smaller numbers are more frequently used. In this encoding method, the most significant bit of each byte has a special meaning. If the most significant bit is 1, it indicates that the subsequent bytes are also a part of the number. If the most significant bit is 0, it indicates that the number ends. That is, only seven bits of each byte are valid data, and the most significant bit is a flag, indicating whether the current byte is the last byte. If the bit width of a number is greater than 7 bits, multiple bytes are required to represent the number. In addition, the number is in big-endian order, and the least significant digit is the first byte.

GPB Parsing Example of the Proto File Header

The following figure shows a GPB parsing example of a proto file.

Figure 19-23 .proto file header packet



- 0A 03 49 41 53: field index 1, string, node_id_str = IAS
- 12 05 73 75 62 30 32: field index 2, string, subscription_id_str = sub02
- 1A 22 68 75... 63 6F 72 64: field index 3, string, sensor_path = huawei-ueip:kpi-records/kpi-recor
- 20 3D: field index 4, uint, collection_id = 0x3d
- 28 F2 D0 A3 BD D2 1B: field index 5, uint, collection_start_time

 Copyright © Huawei Technologies Co., Ltd.

- 30 E8 CD A3 BD D2 1B: field index 6, uint, msg_timestamp
- 3A CA 09: field index 7, variable length, length = 0xCA09=0x4ca = 1226. The data_gpb field is a variable of the TelemetryGPBTable structure and its length is uncertain at this layer. Therefore, the length of this field is variable.
- 0A 97 03: field index 1, variable length, length = 0x9703 = 0x197 = 407. Note that the field index changes to 1 again because the first variable row of the corresponding structure TelemetryGPBTable is repeated before the row defined by proto, indicating that the row is an array. Therefore, it indicates the total data length of the row[] array.
- 08 F2 D0 A3 BD D2 1B: field index 1, int, TelemetryRowGPB.timestamp. The field index is 1, which is the first field of the TelemetryRowGPB structure.
- 5A 8D 03: field index 11, variable length, length=0x8D03=0x18D=397, TelemetryRowGPB.content. This field is followed by the reported service data. It is defined by bytes and is variable-length data. The length is the total number of bytes of the subsequent service data.
- The subsequent field content does not belong to the Telemetry header, but belongs to the service data Telemetry.
- 0A 8A 03: field index 1, variable length, length=0x8A03=0x18A=394, corresponding to the outermost repeated KpiRecord kpi_record of proto of the service data. Note that this is a variable definition with repeated annotations. It is also an array. Therefore, 394 is the length of the kpi_record[] array.

19.5.6 Telemetry Configurations

The control module of the collector uses NETCONF to deliver data collection configurations based on the configured YANG data model.

NOTE

Before configuring Telemetry, use the Telemetry packet parsing tool to start server monitoring, start the NETCONF server, and import the YANG file.

Perform the following steps to configure Telemetry static subscription:

1. Enable NETCONF.
2. Configure the data to be sampled.
3. Configure a destination collector.
4. Create a subscription.

19.5.6.1 Enabling NETCONF

Procedure

1. Run the **sysman server source netconf** command to bind a source port to the NETCONF service protocol.
2. Run the **sysman service netconf enable** command to enable the NETCONF service and create a NETCONF user.

Example

- Bind the NETCONF service protocol to a specified source port. Set the port type to vlanif and the port index to 300.

 Copyright © Huawei Technologies Co., Ltd.

- Enable NETCONF and set the user name and password. Here, V100R021C0x is used as an example. For other versions, the actual display prevails.

```
huawei#sysman server source netconf vlanif 300
huawei(config)#sysman service netconf enable
Please enter the username and password to create a netconf administrator
user
User Name(length<6,15>):Test-
user1
User
Password(length<12,128>):
Confirm
Password(length<12,128>):
Adding user successfully
```

19.5.6.2 Configuring the Data to Be Sampled

Create a sampling sensor group and specify a sampling path.

Table 19-11 Elements for configuring data to be sampled

Element	Description	Parameter
sensor-group-id	Indicates the ID of a sampling sensor group.	The value is a string of 1–63 characters, including uppercase letters, lowercase letters, and digits (0–9).
path	Indicates the sampling path.	The value is a string of 1–254 characters.

Set sensor-group-id to sen01 and path to huawei-gpon-pm-ont-line-quality:gpon-pm-ont-line-qualitys/gpon-pm-ont-line-quality. The following is a YANG configuration example. If the configuration is successful, OK is returned.

- Request example:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="1581665858050">
 <edit-config>
  <target>
   <running />
  </target>
  <config xmlns:xc="urn:ietf:params:xml:ns:netconf:base:1.0">
   <telemetry-system xmlns="http://openconfig.net/yang/telemetry">
    <sensor-groups>
     <sensor-group>
      <sensor-group-id>sen01</sensor-group-id>
      <config>
       <sensor-group-id>sen01</sensor-group-id>
      </config>
      <sensor-paths>
       <sensor-path xc:operation="create">
        <path>huawei-gpon-pm-ont-line-quality:gpon-pm-ont-line-qualitys/gpon-pm-ont-line-quality</path>
        <config>
         <path>huawei-gpon-pm-ont-line-quality:gpon-pm-ont-line-qualitys/gpon-pm-ont-line-quality</path>
        </config>
       </sensor-path>
      </sensor-paths>
     </sensor-group>
```

 Copyright © Huawei Technologies Co., Ltd.

```
      </sensor-groups>
    </telemetry-system>
  </config>
 </edit-config>
</rpc>
```

- Response example:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="1581665858050">
 <ok />
</rpc-reply>
```

Query the configured data to be sampled. The following is an example:

- Request example:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="1581665858121">
 <get>
  <filter type="subtree">
   <telemetry-system xmlns="http://openconfig.net/yang/telemetry" xmlns:hw-telemetry-
ext="urn:huawei:yang:huawei-openconfig-telemetry-ext">
    <sensor-groups>
     <sensor-group>
      <sensor-paths>
       <sensor-path />
      </sensor-paths>
     </sensor-group>
    </sensor-groups>
   </telemetry-system>
  </filter>
 </get>
</rpc>
```

- Response example:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="1581665858121">
 <data>
  <telemetry-system xmlns="http://openconfig.net/yang/telemetry">
   <sensor-groups>
    <sensor-group>
     <sensor-group-id>sen01</sensor-group-id>
     <sensor-paths>
      <sensor-path>
       <path>huawei-gpon-pm-ont-line-quality:gpon-pm-ont-line-qualitys/gpon-pm-ont-line-
quality</path>
       <config>
        <path>huawei-gpon-pm-ont-line-quality:gpon-pm-ont-line-qualitys/gpon-pm-ont-line-
quality</path>
       </config>
       <state>
        <path>huawei-gpon-pm-ont-line-quality:gpon-pm-ont-line-qualitys/gpon-pm-ont-line-
quality</path>
       </state>
      </sensor-path>
     </sensor-paths>
    </sensor-group>
   </sensor-groups>
  </telemetry-system>
 </data>
</rpc-reply>
```

19.5.6.3 Configuring a Destination Collector

Create a destination collector group and specify the destination collector to which the sampled data is sent.

 Copyright © Huawei Technologies Co., Ltd.

Table 19-12 Elements for configuring a destination collector

Element	Description	Parameter
group-id	Indicates the ID of a destination collector group.	The value is a string of 1–63 characters, including uppercase letters, lowercase letters, and digits (0–9).
destination-address	Indicates a destination IP address.	Loopback IP addresses, all-0 IP addresses, all-f IP addresses, and IPv6 addresses are not supported.
destination-port	Indicates a destination port.	The value is an integer ranging from 1 to 65535.

Set group-id to dest01, destination-address to 192.168.1.1, and destination-port to 20001. The following is a YANG configuration example. If the configuration is successful, OK is returned.

- Request example:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="1579172235104">
 <edit-config>
  <target>
   <running />
  </target>
  <config xmlns:xc="urn:ietf:params:xml:ns:netconf:base:1.0">
   <telemetry-system xmlns="http://openconfig.net/yang/telemetry">
    <destination-groups>
     <destination-group>
      <group-id>dest01</group-id>
      <config>
       <group-id>dest01</group-id>
      </config>
      <destinations>
       <destination xc:operation="create">
        <destination-address>192.168.1.1</destination-address>
        <destination-port>20001</destination-port>
        <config>
         <destination-address>192.168.1.1</destination-address>
         <destination-port>20001</destination-port>
        </config>
       </destination>
      </destinations>
     </destination-group>
    </destination-groups>
   </telemetry-system>
  </config>
 </edit-config>
</rpc>
```

- Response example:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="1579172235104">
 <ok />
</rpc-reply>
```

 Copyright © Huawei Technologies Co., Ltd.

Query information about the configured destination collector. The following is an example:

- Request example:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="1581665858129">
 <get>
  <filter type="subtree">
   <telemetry-system xmlns="http://openconfig.net/yang/telemetry">
    <destination-groups>
     <destination-group>
      <destinations>
       <destination />
      </destinations>
     </destination-group>
    </destination-groups>
   </telemetry-system>
  </filter>
 </get>
</rpc>
```

- Response example:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="1581665858129">
 <data>
  <telemetry-system xmlns="http://openconfig.net/yang/telemetry">
   <destination-groups>
    <destination-group>
     <group-id>dest01</group-id>
     <destinations>
      <destination>
       <destination-address>192.168.1.1</destination-address>
       <destination-port>20001</destination-port>
       <config>
        <destination-address>192.168.1.1</destination-address>
        <destination-port>20001</destination-port>
       </config>
       <state>
        <destination-address>192.168.1.1</destination-address>
        <destination-port>20001</destination-port>
       </state>
      </destination>
     </destinations>
    </destination-group>
   </destination-groups>
  </telemetry-system>
 </data>
</rpc-reply>
```

19.5.6.4 Creating a Subscription

Create a subscription, associate the configured destination group with the sampling sensor group, and establish a data sending channel to complete data sending. Data can be sent through UDP.

Table 19-13 Elements for creating a subscription

Element	Description	Parameter
subscription-name	Indicates the subscription name.	The value is a string of 1–63 characters, including uppercase letters, lowercase letters, and digits (0–9).

 Copyright © Huawei Technologies Co., Ltd.

Element	Description	Parameter
local-source-address	Indicates a source IP address.	The value is an IPv4 address.
protocol	Indicates the protocol.	The STREAM_UDP type is supported.
encoding	Indicates the encoding format.	The ENC_PROTO3 type is supported.
local-source-port	Indicates a source port.	The value is an integer ranging from 1 to 65535.
sensor-group	Indicates the sampling sensor group to be associated with the subscription.	The value is a string of 1–63 characters, including uppercase letters, lowercase letters, and digits (0–9). The sensor-group must have been created.
sample-interval	Indicates the sampling interval.	The value is an integer ranging from 1 to 86400000, in milliseconds.
group-id	Indicates the destination collector group to be associated with the subscription.	The value is a string of 1–63 characters, including uppercase letters, lowercase letters, and digits (0–9). The destination-group must have been created.

Set subscription-name to sub01, local-source-address to 10.71.214.35, protocol to STREAM_UDP, encoding to ENC_PROTO3, local-source-port to 55, sensor-group to sen01, sample-interval to 60000, and group-id to dest01. The following is a YANG configuration example. If the configuration is successful, OK is returned.

- Request example:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="1579172235116">
 <edit-config>
  <target>
   <running />
  </target>
  <config xmlns:xc="urn:ietf:params:xml:ns:netconf:base:1.0">
   <telemetry-system xmlns="http://openconfig.net/yang/telemetry" xmlns:hw-telemetry-
ext="urn:huawei:yang:huawei-openconfig-telemetry-ext">
    <subscriptions>
     <persistent>
      <subscription xc:operation="create">
       <subscription-name>sub01</subscription-name>
       <config>
        <subscription-name>sub01</subscription-name>
        <local-source-address>10.71.214.35</local-source-address>
```

 Copyright © Huawei Technologies Co., Ltd.

```
            <protocol>STREAM_UDP</protocol>
            <encoding>ENC_PROTO3</encoding>
            <hw-telemetry-ext:local-source-port>55</hw-telemetry-ext:local-source-port>
          </config>
          <sensor-profiles>
            <sensor-profile>
              <sensor-group>sen01</sensor-group>
              <config>
                <sensor-group>sen01</sensor-group>
                <sample-interval>60000</sample-interval>
              </config>
            </sensor-profile>
          </sensor-profiles>
          <destination-groups>
            <destination-group>
              <group-id>dest01</group-id>
              <config>
                <group-id>dest01</group-id>
              </config>
            </destination-group>
          </destination-groups>
        </subscription>
      </persistent>
    </subscriptions>
  </telemetry-system>
 </config>
</edit-config>
</rpc>
```

- Response example:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="1579172235116">
 <ok />
</rpc-reply>
```

Query the created subscription information. The following is an example:

- Request example:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="1579172235130">
 <get>
  <filter type="subtree">
   <telemetry-system xmlns="http://openconfig.net/yang/telemetry">
    <subscriptions>
     <persistent>
      <subscription />
     </persistent>
    </subscriptions>
   </telemetry-system>
  </filter>
 </get>
</rpc>
```

- Response example:

```
<?xml version="1.0" encoding="UTF-8"?>
<rpc-reply xmlns="urn:ietf:params:xml:ns:netconf:base:1.0" message-id="1579172235130">
 <data>
  <telemetry-system xmlns="http://openconfig.net/yang/telemetry">
   <subscriptions>
    <persistent>
     <subscription>
      <subscription-name>sub01</subscription-name>
      <config>
       <subscription-name>sub01</subscription-name>
       <local-source-address>10.71.214.35</local-source-address>
       <protocol xmlns:hw-telemetry-ext="urn:huawei:yang:huawei-openconfig-telemetry-ext">hw-
telemetry-ext:STREAM_UDP</protocol>
       <encoding xmlns:oc-telemetry-types="http://openconfig.net/yang/telemetry-types">oc-
telemetry-types:ENC_PROTO3</encoding>
       <local-source-port xmlns="urn:huawei:yang:huawei-openconfig-telemetry-ext">55</local-
```

 Copyright © Huawei Technologies Co., Ltd.

```
source-port>
          </config>
          <state>
            <subscription-name>sub01</subscription-name>
            <local-source-address>10.71.214.35</local-source-address>
            <protocol xmlns:hw-telemetry-ext="urn:huawei:yang:huawei-openconfig-telemetry-ext">hw-
telemetry-ext:STREAM_UDP</protocol>
            <encoding xmlns:oc-telemetry-types="http://openconfig.net/yang/telemetry-types">oc-
telemetry-types:ENC_PROTO3</encoding>
            <local-source-port xmlns="urn:huawei:yang:huawei-openconfig-telemetry-ext">55</local-
source-port>
          </state>
          <sensor-profiles>
            <sensor-profile>
              <sensor-group>sen01</sensor-group>
              <config>
                <sensor-group>sen01</sensor-group>
                <sample-interval>60000</sample-interval>
              </config>
              <state>
                <sensor-group>sen01</sensor-group>
                <sample-interval>60000</sample-interval>
              </state>
            </sensor-profile>
          </sensor-profiles>
          <destination-groups>
            <destination-group>
              <group-id>dest01</group-id>
              <config>
                <group-id>dest01</group-id>
              </config>
              <state>
                <group-id>dest01</group-id>
              </state>
            </destination-group>
          </destination-groups>
        </subscription>
      </persistent>
    </subscriptions>
  </telemetry-system>
 </data>
</rpc-reply>
```

19.6 Inband Management VPN

Inband management VPN is a means by which carriers use the virtual private network (VPN) to manage and maintain devices and the management protocol on the device can use virtual routers for route forwarding.

19.6.1 Introduction to Inband Management VPN

Definition

In inband management VPN (Virtual Private Network), associated inband management protocols on the device support the specified VPN instances so that management packets can be received and forwarded using multiple virtual routes. In this way, carriers can use the private network IP address to remotely manage and maintain devices. This method saves public network IP addresses and isolates the management network from the public network.

To achieve inband management VPN, both the inband management server and client must be able to receive the connection requests and data packets from VPN.

 Copyright © Huawei Technologies Co., Ltd.

Constraints

- The outband management interface belongs to the public network but not VPN. Therefore, only the inband interface but not the outband management interface supports VPN management.
- The servers that can receive VPN requests include:
  - Telnet server
  - SSH server
  - SNMP AGENT
  - TRACE server

  NOTE

  SSH server is recommended.
- The clients that can receive VPN requests include:
  - FTP client
  - TFTP client
  - SFTP client
  - SNMP TRAP
  - SYSLOG
  - Telnet client
  - SSH client

  NOTE

  SFTP client is recommended.

19.6.2 Principles to Inband Management VPN

Basic Concepts

Virtual private network (VPN) is a network technology for encapsulating or encrypting private data and then transmitting the data over the public network. With this technology, the security level of the private network can be provided for the transmitted data and a private network can be constructed based on the public network. VPN is a logical private network that provides the functions of the private network. The network itself, however, is not an independent physical network. In the IP bearer network, VPN is an important measure for logically isolating services, preventing attacks, and helping implement QoS control.

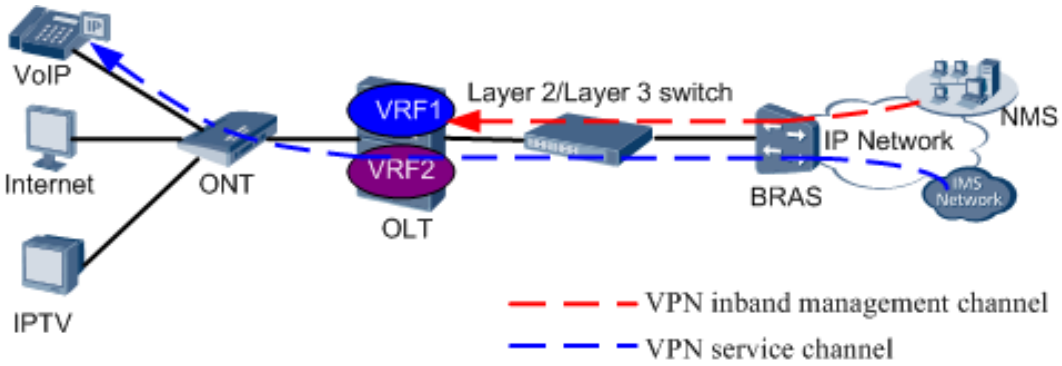
A VPN instance is also called a VPN routing and forwarding table (VRF). Each router is logically divided into multiple virtual routers, that is, multiple VRFs. Each VRF corresponds to a VPN and has its own routing table, forwarding table, and interface. In other words, a router shared by various VPNs is simulated as multiple private routers, thereby isolating VPN routes. Devices that are grouped into a private route exchange routing information of only the private route. After VPN maintenance and management instances are configured, the server end protocols of the device, that is, telnet server ends (SSH and CLI), SNMP, and TRACE, accept connection requests of only the specified management and maintenance VPN networks and do not accept those of other VPNs or of the public network.

 Copyright © Huawei Technologies Co., Ltd.

Inband Management VPN

Inband management VPN uses the VRF function to add the remote NMS and the OLT to the same VPN while on the OLT classifying the management address and VoIP address to different VRFs. In this way, carriers can use the private network IP address to remotely manage and maintain devices. This method saves public network IP addresses and isolates the management network from the public network.

Figure 19-24 Network diagram of inband management VPN



In **Figure 19-24**, two VRFs (VRF1 and VRF2) are defined on the OLT. VRF2 is the VoIP service channel and VRF1 is the VPN inband management channel.

On the OLT, if a system-level VPN instance for management and maintenance is configured, all related management protocols by default use this VPN instance to connect to or send data to the remote server. In addition, if a user specifies the VPN instance of a single server when configuring the trap destination server or running the telnet command, these two management protocols (for trap and telnet) do not use the system-level VPN instance but instead use the specified VPN instance to connect to the remote server.

19.7 Remote Software Commissioning (Ethernet)

This section describes the implementation principles and configuration of remote software commissioning using Ethernet upstream transmission.

19.7.1 Introduction

During the deployment of the multi-dwelling unit (MDU) in the Ethernet upstream networking or the deployment of the OLT, you need to configure management channel parameters after installing the device so that the device can be managed. To avoid onsite commissioning and reduce the engineering deployment cost, the MDU and OLT provide the remote software commissioning function for Ethernet upstream transmission. After a device is powered on, management data and service data can be automatically configured.

Remote software commissioning for Ethernet upstream transmission can be implemented based on Dynamic Host Configuration Protocol (DHCP) or neighbor automatic communication (NAC).

 Copyright © Huawei Technologies Co., Ltd.

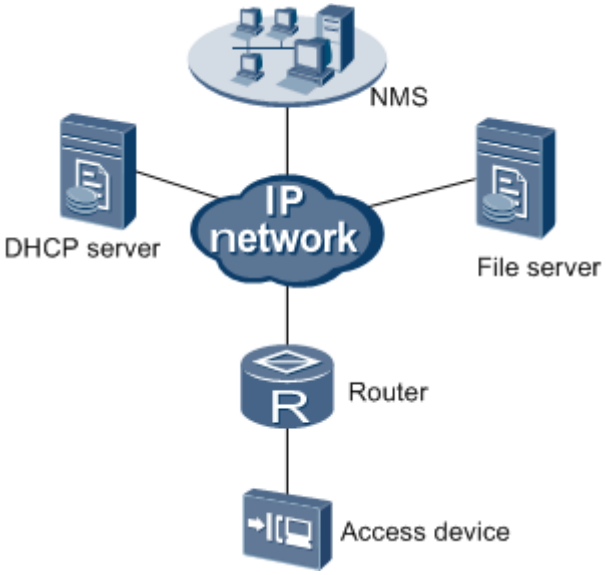
NOTE

Run the **sysman server source** command to bind a specified source port to the SNMP/SSH/NETCONF service protocol first in V100R020C00 and later versions.

DHCP-based Remote Software Commissioning for Ethernet Upstream Transmission

Figure 19-25 shows the networking diagram for this solution.

Figure 19-25 Networking of DHCP-based remote software commissioning for Ethernet upstream transmission



1. Bind the IP address and MAC address of the access device (MDU or OLT) according to the network planning, and configure the data on the DHCP server.
2. After the access device is powered on, it automatically sends a DHCP request to the DHCP server. The DHCP server allocates an IP address to the access device based on the mapping between the IP address and the MAC address

NOTICE

If you log in to the device when the DHCP remote software commissioning function is enabled, the remote software commissioning process may be interrupted. To restart the remote software commissioning function, clear the database and reset the device.

3. The access device requests the policy file from the FTP server and loads the policy file for remote commissioning.

 Copyright © Huawei Technologies Co., Ltd.

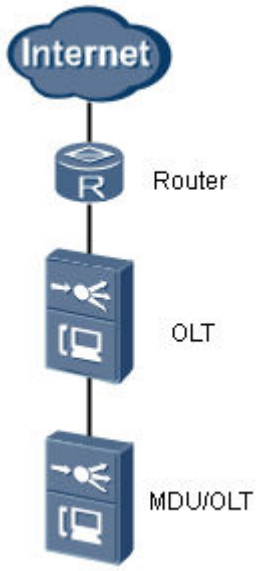
NOTICE

- DHCP-based Remote Software Commissioning packets are not encrypted. Ensure that the DHCP-based Remote Software Commissioning networking environment is secure. NETCONF ZTP software commissioning-free mode that supports identity authentication and encrypted transmission is recommended in V100R020C10 and later versions.
- Before V100R020C10 version, the SFTP protocol is not supported in DHCP-based Remote Software Commissioning. FTP and TFTP have security risks due to protocol restrictions, and can be used only in a secure environment.
- SFTP is recommended in DHCP-based Remote Software Commissioning with a higher security in V100R020C10 and later versions.

NAC-based Remote Software Commissioning for Ethernet Upstream Transmission

Figure2 Networking of NAC-based remote software commissioning for Ethernet upstream transmission show the networking diagrams for this solution.

Figure 19-26 Networking of NAC-based remote software commissioning for Ethernet upstream transmission



The NAC-based remote software commissioning solution for Ethernet upstream transmission applies two scenarios: The database of the cascaded device is empty, or the database of the cascaded device is not empty.

- If the database of the cascaded device is empty, the NAC protocol is used to implement remote software commissioning for Ethernet upstream transmission. In this scenario, the upper layer of the Internet can also be connected to the NMS to configure management channel parameters.
- If the database of the cascaded device is not empty, you need to log in to the cascaded device on the CLI of the transparent channel and perform

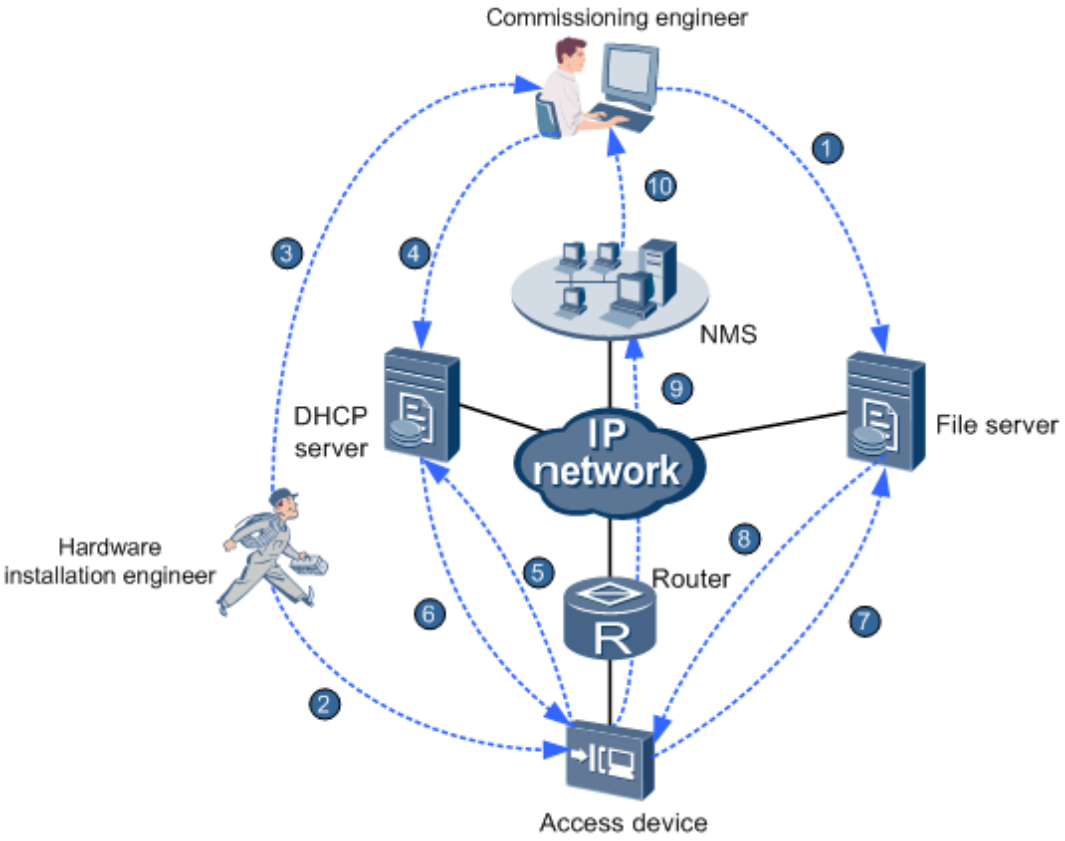
 Copyright © Huawei Technologies Co., Ltd.

operations to implement the remote software commissioning for Ethernet upstream transmission.

19.7.2 Principles (Based on DHCP)

Figure 19-27 shows the implementation of DHCP-based remote software commissioning for Ethernet upstream transmission.

Figure 19-27 Implementation of DHCP-based remote software commissioning for Ethernet upstream transmission



NOTE

DHCP-based remote software commissioning for Ethernet upstream transmission is configured on the upper-layer devices instead of on access devices.

Before R019C20 version, the DHCP packets sent to the upper-layer device are untagged. A port default VLAN ID (PVID) must be configured on the upper-layer device to transparently transmit these DHCP packets; in R019C20 and later versions, the access device updates the VLAN (VLAN 1 to VLAN 16) and upstream ports one by one and attempts to obtain an IP address through DHCP.

The process is as follows:

1. A commissioning engineer prepares the pre-configuration policy file and policy description file and uploads them to the FTP server.

   For a policy file:

   - If different devices require different policy files, the policy files can be named in the [IP]_stFile.txt or [MAC]_stFile.txt format. If the policy file is

 Copyright © Huawei Technologies Co., Ltd.

in the [IP]_stFile.txt format, ensure that the IP address of the device does not conflict with those of other devices. You can bind the IP address of a device to a MAC address to avoid IP address conflict.

- If devices of same type use the same policy file, the name of the configuration file can be a common string.
  - IP indicates the IP address planned for a device. For example, if the IP address of a device is 192.168.10.10, the policy file of the device is named 192.168.10.10_stFile.txt.
  - MAC indicates the actual MAC address of a device. For example, if the MAC address of a device is 00E0FC11ABEE, the policy file of the device is named 00-E0-FC-11-AB-EE_stFile.txt.
  - The policy file name is a common character string, for example, OLTxxxstFile.txt *(example)*. All devices of the same type use the policy file.

For a policy description file:

- The name of a policy description file must be in the format of xxx.ini. The file content includes the path for storing the policy file and the name of the policy file.
- The format and content of a policy description file must be correct. Otherwise, the remote software commissioning flow may be interrupted. If this happens, operators must remotely log in to the device using the IP address allocated by the DHCP server and reset the device. Alternatively, operators can reset the device onsite. Then, the remote software commissioning flow can be restarted.
- One site requires only one policy description file.
- For example, the policy file for a site is named by an IP address. The content of the policy description file is as follows (the following is only an example, and the file name must be consistent with the actual plan):

```
[OLT]                                   //Device type
TargetVersion=OLT Vx00RxxxCxx           //Device version
[HxxxXXXX]                              //Control board type (example)
PacketFile="Dir\xxxx_packetfile.bin"     //Package file to be loaded to the control board (Example)
CfgFile="Dir\[IP]_stFile.txt"            //Policy file to be loaded to the control board, in the IP_stFile.txt naming format
```

2. A hardware installation engineer obtains the device, transports the device from the warehouse to a site, and installs, connects, and powers on the device.

NOTICE

Do not remove or insert a board when the device is starting. Otherwise, the device data may be abnormal.

3. The hardware installation engineer records the MAC address and site information of the access device and reports the information to the commissioning engineer.
4. The commissioning engineer creates a mapping between the MAC address, management IP address, and physical location of the access device, binds the

 Copyright © Huawei Technologies Co., Ltd.

IP and MAC addresses of the device, and configures the device data on the DHCP server.

5. After the upstream optical path of the access device is normal (the LINK indicator is steady on), the access device automatically initiates a DHCP request to the DHCP server.
6. The DHCP server allocates the IP address to the access device based on the IP address configuration, writes the FTP information (including the FTP server address, FTP user name, password, and policy description file name) in an option packet, and sends the option packet to the access device.
7. After receiving the Option packet from the DHCP server, the access device configures the IP address and gateway and sends a request to the FTP server.
8. The FTP server delivers a policy file to the access device.
9. The access device downloads the policy file from the FTP server and automatically loads the file to complete automatic device configuration. The configuration takes effect after the device restarts. At the same time, a trap is reported to the NMS, indicating that an access device goes online.

   NOTE

   If the loading fails, the system automatically retries for about 24 hours. If the loading still fails, the loading is stopped.
10. The commissioning engineer receives the access device online trap on the NMS.

    When an access device online trap is received, a network management channel has been created, and the access device can be remotely managed on the NMS.

19.7.3 Configuring NAC-based Remote Software Commissioning for Ethernet Upstream Transmission

When devices are cascaded to an OLT that supports the NAC protocol, management channels to these devices can be automatically enabled without the need for a commissioning engineers to go onsite.

Prerequisites

Management parameters, including the inband management VLAN ID, management IP address, and SNMP profile parameters, have been configured on the cascading device for cascaded devices. (SNMP profile parameters are required if the cascaded devices needs to be managed on the NMS.)

Context

This section uses the scenario of connecting an MDU to an existing OLT as an example to describe the configuration process. For the scenario where an OLT is cascaded to an OLT, the configuration process is the same as that used in scenario where an MDU is cascaded to an OLT, except that the former scenario does not support automatic deployment of policy files.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

- The MDU does not save the management parameters delivered through NAC. To save the parameters, run the **save** command. After saving the parameters, the NAC function will be disabled automatically.

In V100R021C00 and later versions, you are advised to run the **nac authentication-mode** command to configure the authentication mode for the NAC channel. By default, the authentication mode of an NAC channel is certificate authentication. If the authentication mode of the NAC channel is certificate authentication, you need to bind the PKI domain to the NAC channel for normal authentication. You can run the **nac bind pki-domain** command to bind a PKI domain.

NOTICE

Huawei devices are delivered with preconfigured certificates which are applied for and managed using the Huawei PKI platform. To improve system security, you are advised to replace the preconfigured certificates and keys with your own certificates and keys and periodically update the certificates and keys.

Pay attention to the validity period of the certificates. Generally, the start time of a certificate is the time when the certificate is issued, and the end time is the time when the certificate expires. You can run the **display pki certificate** command to query the validity period of a certificate.

When a certificate is about to expire or has expired, the system reports an alarm indicating that the certificate is about to expire or has expired, prompting users to promptly update the certificate.

Device certificates can be updated using the following methods:

- **Method 1: Apply for a certificate online using CMPv2.**
  a. Run the **rsa pki local-key-pair create** command to create a local RSA (Ron Rivest, Adi Shamirh, Leonard Adleman) key pair.
  b. Run the **pki entity** command to create a PKI entity.
  c. Run the **pki domain** command to create a PKI domain and run the **pki import-certificate** command to import a certificate file.
  d. Run the **ssl policy** command to create an SSL policy and run the **pki-domain** command to bind a PKI domain.
  e. Run the **pki cmp session** command to create a Certificate Management Protocol (CMP) session and enter CMP session mode.
  f. Run the **cmp request** command to configure a local certificate, CA certificate, and URL of the CMP server.
  g. Run the **pki domain** command to enter PKI domain mode, and run the **certificate request** command to specify the local RSA key pair, PKI entity information, and CMP session information for certificate application.
  h. In PKI domain mode, run the **cmp initial-request** command to apply for a certificate online.
- **Method 2: Apply for a certificate offline and import it to the device.**

 Copyright © Huawei Technologies Co., Ltd.

a. Apply for a certificate outside the device.

b. Run the **load pki certificate** command to load the obtained certificate to the device.

Procedure

- The MDU database is empty.

  a. Enable NAC on the OLT.

  The OLT port connected to the MDU is 0/2/0.

  ```
  huawei(config)#nac enable master security 0/2 0
  ```

  NOTE

  When enabling NAC on the OLT, you are advised to select **security** to enable the security mode. Then, the master node encrypts remote software commissioning parameters before sending them to a slave node. This improves parameter security.

  b. Pre-configure remote software commissioning parameters for the MDU on the OLT and check whether the pre-configuration is successful.

  The MDU inband management VLAN ID is 8, the management IP address is 192.168.50.2, the IP address of the gateway is 192.168.50.254, and the SNMP profile ID is 64. These parameters can be configured in two ways:

  NOTE

  When both NAC and RN configuration commands are executed, the latter overrides the previous one.

  Configure and query parameters using NAC commands:

  ```
  huawei(config)#nac config master 0/2/0 slave vlan 8 priority 3 ip-address  192.168.50.2 mask 255.255.255.0 gateway 192.168.50.254 snmp-profile-id 64
  ```

  ```
  huawei(config)#display nac configuration master
  { <cr>|frameid/slotid/portid<S><Length 5-18>|frameid/slotid<S><Length 3-15> }:

   Command:
         display nac configuration master
   ------------------------------------------------------------------------------
   Frame ID/Slot ID/Port ID :  0/2/0
   VLAN                     :  8
   VLAN priority            :  3
   IP address               :  192.168.50.2
   Mask                     :  255.255.255.0
   Gateway                  :  192.168.50.254
   Snmp-profile-id          :  64
   Snmp-profile-name        :  snmp-profile_64
   ------------------------------------------------------------------------------
   Total: 1
  ```

  Configure and query parameters using RN commands:

  ```
  huawei(config)#rn ipconfig 0/2/0 ip-address 192.168.50.2 mask 255.255.255.0 gateway 192.168.50.254 vlan 8 priority 3
  huawei(config)#rn snmp-config 0/2/0 snmp-index 0 profile-id 64
  ```

  ```
  huawei(config)#display rn info 0/2/0
   Command:
       display rn info 0/2/0
   ------------------------------------------------------------------------------
   F/S/P              :  0/2/0
   Equipment ID       :  XXXXXX  //"XXXXXX" is the MDU name.
   Uplink type        :  ETH
   MAC/SN             :  xxxx-xxxx-xxxx
  ```

 Copyright © Huawei Technologies Co., Ltd.

```
Run state          : Offline
Config state       : -
IP                 : 192.168.50.2
Subnet mask        : 255.255.255.0
Gateway            : 192.168.50.254
Manage VLAN        : 8
Manage priority    : 3
Deploy profile ID  : -
Deploy profile name : -
-------------------------------------------------------------------
Index  SNMP        SNMP                          Effective
       Profile ID  Profile name                  Flag
-------------------------------------------------------------------
0      64          snmp-profile_64               No
1      -           -                             -
-------------------------------------------------------------------
-------------------------------------------------------------------------------
```

c. Configure an automatic deployment policy profile (applicable only to the scenario when the OLT is cascaded with an MDU).

   The profile name is deploy-profile_1, the policy file name of the MDU is stfile, the IP address is 10.10.10.10, the file transfer protocol is FTP, the user name is user1, and the password is user123 (the system does not display the entered password).

```
huawei(config)#rn deploy-profile add  profile-id 1 profile-name deploy-profile_1 filename stfile ip 10.10.10.10 ftp user user1
```

d. Bind the automatic deployment policy profile to the MDU (applicable only to the scenario when the OLT is cascaded with an MDU).

   Bind the profile based on ports:

```
huawei(config)#rn deploy-config 0/2/0 profile-id 1
```

   Bind the profile based on device types:

```
huawei(config)#rn deploy-policy MAXXXX profile-id 1 //Change MAXXXX to the actual MDU type.
```

   NOTE

   The port-based profile binding takes precedence over the device type-based profile binding.

e. After the MDU is powered on, it exchanges data with the OLT to obtain the MDU device management parameters and service configuration parameters. Wait for 2 minutes. Then, run the **ssh** command to remotely log in to the MDU.

f. Query the devices connected to the OLT cascading port.

   The query can be performed in two ways:

   Using NAC commands:

```
huawei(config)#display nac slave info detail
-------------------------------------------------------------------------------
NAC master port          : 0/2/0
NAC auto-find fsm state  : Established
NAC auto-find result     : Found
Multiple slave node exist : No
Slave node device type   : XXXXXX  //"XXXXXX" is the MDU name.
Slave node mac-address   : 00-E0-FC-01-00-52
NAC slave port           : 0/0/1
NAC config fsm state     : INIT
NAC config result        : Added successfully
NAC config fail reason   : -
-------------------------------------------------------------------------------
```

   Using RN commands:

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#display rn info 0/2/0
 Command:
      display rn info 0/2/0
-------------------------------------------------------------------------------
F/S/P             :  0/2/0
Equipment ID      :  XXXXXX  //"XXXXXX" is the MDU name.
Uplink type       :  ETH
MAC/SN            :  xxxx-xxxx-xxxx
Run state         :  Online
Config state      :  Normal
IP                :  192.168.50.2
Subnet mask       :  255.255.255.0
Gateway           :  192.168.50.254
Manage VLAN       :  8
Manage priority   :  3
Deploy profile ID :  1
Deploy profile name :  deploy-profile_1
-----------------------------------------------------------------
Index  SNMP       SNMP                        Effective
       Profile ID  Profile name                Flag
-----------------------------------------------------------------
0      64         snmp-profile_64              Yes
1      -          -                            -
-----------------------------------------------------------------
-------------------------------------------------------------------------------
```

g. (Optional) Disable NAC on the OLT.

Perform this operation after the MDU obtains the device management parameters. This prevents the OLT from delivering the planned data to a substitute MDU.

```
huawei(config)#nac master 0/2 0 lock
```

- The MDU database is not empty.

NOTICE

The packets transmitted through a transparent channel are not encrypted and involves risks. Exercise caution when performing this operation.

The following describes only the configuration process if the MDU database must be erased. The configuration is simple if the MDU database does not need to be erased because you only need to perform the operations in **b** to log in to the MDU and configure device management parameters on the MDU.

NOTE

The transparent channel is enabled on the MDU by default. To ensure system security or if the transparent channel is not used, run the **transparent remote disable** command to disable it.

a. Perform steps **a**–**d**.

b. After the MDU is powered on, enable the transparent channel on the OLT port, and log in to the MDU.

The OLT port connected to the MDU is 0/2/0.

```
huawei(config)#diagnose
huawei(diagnose)%%transparent on 0/2/0
>>User name:huawei  //Enter the user name of the MDU.
>>User password:    //Enter the user password.
```

 Copyright © Huawei Technologies Co., Ltd.

```
Huawei Integrated Access Software (huawei).
Copyright(C) Huawei Technologies Co., Ltd. 2002-2013. All rights reserved.

------------------------------------------------------------------------------
User last login information:
------------------------------------------------------------------------------
Access Type : Serial
IP-Address  : --
Login  Time : 2013-07-21 01:04:17+08:00
Logout Time : 2013-07-21 05:19:35+08:00
------------------------------------------------------------------------------

User root has used a default password. Change the password in time.
```

c. Erase the MDU database.

NOTE

The default command mode of **erase flash data** varies with the software version. The following configuration is for reference only.

```
huawei>enable
huawei#erase flash data
  This command will clear the active board data that has been loaded or saved
                                                    Please remember to backup the system
configuration data
  Are you sure to continue? (y/n)[n]:y
  Command executes successfully, and will take effect after active board
rebooted
```

d. Reset the MDU.

```
huawei#reboot system
  Please check whether data has saved, the unsaved data will lose if reboot
system, are you sure to reboot system? (y/n)[n]:y
```

e. After the MDU is reset, OLT commands are displayed. Wait for 2 minutes, then check the remote software commissioning result.

The query can be performed in two ways:

Using NAC commands:

```
huawei(diagnose)%%quit
huawei#display nac slave info detail
  ------------------------------------------------------------------------------
  NAC master port          :  0/2/0
  NAC auto-find fsm state   :  Established
  NAC auto-find result      :  Found
  Multiple slave node exist :  No
  Slave node device type    :  XXXXXX  //"XXXXXX" is the MDU name.
  Slave node mac-address    :  00-E0-FC-01-00-52
  NAC slave port           :  0/0/1
  NAC config fsm state      :  INIT
  NAC config result         :  Added successfully
  NAC config fail reason    :  -
  ------------------------------------------------------------------------------
  Total: 1
```

Using RN commands:

```
huawei(config)#display rn info 0/2/0
  Command:
       display rn info 0/2/0
  ------------------------------------------------------------------------------
  F/S/P              :  0/2/0
  Equipment ID       :  XXXXXX  //"XXXXXX" is the MDU name.
  Uplink type        :  ETH
  MAC/SN             :  xxxx-xxxx-xxxx
  Run state          :  Online
  Config state       :  Normal
```

 Copyright © Huawei Technologies Co., Ltd.

```
IP               : 192.168.50.2
Subnet mask      : 255.255.255.0
Gateway          : 192.168.50.254
Manage VLAN      : 8
Manage priority  : 3
Deploy profile ID  : 1
Deploy profile name :  deploy-profile_1
-----------------------------------------------------------------
Index  SNMP       SNMP                          Effective
       Profile ID  Profile name                 Flag
-----------------------------------------------------------------
0      64         snmp-profile_64               Yes
1      -          -                             -
-----------------------------------------------------------------
-------------------------------------------------------------------------------
```

f. Run the **ssh** command to remotely log in to the MDU.

NOTE

The transparent channel commands for logging in to a device are diagnostic commands. After Layer 3 communication is set up, log in to and configure the MDU in remote mode in priority.

----End

19.7.4 Remote Software Commissioning (Ethernet) Reference Standards and Protocols

The following table lists the reference standards and protocols of remote software commissioning using Ethernet upstream transmission.

Standard	Description
IETF RFC 2131	Dynamic Host Configuration Protocol
IETF RFC 1533	DHCP Options and BOOTP Vendor Extensions
IEEE 802.3ah-2004	IEEE Standard for Information technology-Telecommunications and information exchange between systems- Local and metropolitan area networks- Specific requirements Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications Amendment: Media Access Control Parameters, Physical Layers, and Management Parameters for Subscriber Access Networks

19.8 OLT ZTD

With the zero touch deployment (ZTD) feature, OLTs are plug-and-play and can be automatically deployed.

 Copyright © Huawei Technologies Co., Ltd.

19.8.1 Feature Updates

Table 19-14 Feature Updates

Product Version	Change Description
V100R019C2X and later versions	The first version to support this feature

19.8.2 Why the OLT Needs to Support ZTD

When devices are deployed in traditional mode, the following problems exist:

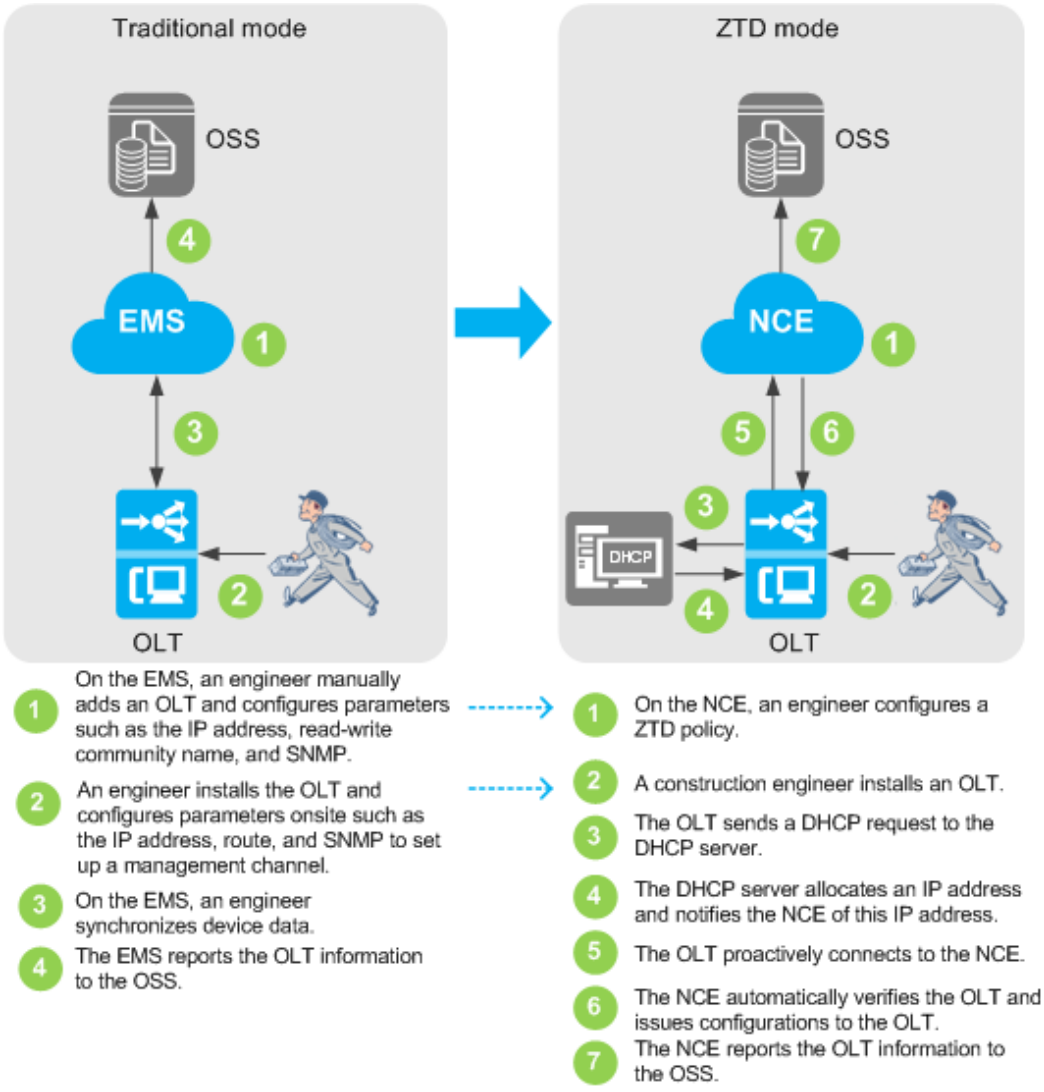
- OLT configuration parameters need to be manually added, which is prone to errors.
- Each OLT needs to be configured separately, which is time-consuming.
- Onsite software commissioning requires a technical foundation for onsite construction personnel and multiple site visits. In this case, plug-and-play is not supported.

As shown in the following figure, a ZTD network can be constructed to implement plug-and-play and automatic deployment of network devices. The ZTD feature improves the service rollout efficiency of network devices and reduces operating costs.

- Unified ZTD configuration policy, verification in advance, zero-error delivery, one-time configuration, and process-based running
- Automatic setup of a management channel without the need of secondary site visits, improving the OLT deployment efficiency and reducing the workload of installation and maintenance personnel

 Copyright © Huawei Technologies Co., Ltd.

Figure 19-28 Comparison between the traditional configuration mode and the ZTD configuration mode



The OLT works with the Network Cloud Engine (NCE) to implement the plug-and-play function. The OLT supports the NETCONF CALL HOME feature (described in RFC 8071). After an OLT is connected to the NCE, the NCE reports the OLT information to the upper-layer equipment serial number (ESN) whitelist system for authentication. An NE supports plug-and-play after it passes the whitelist authentication.

NOTE

After the CALL HOME service is enabled, the system automatically creates a default NETCONF administrator (user name: ncroot; password: NFRoot.123).

19.8.3 OLT ZTD Application Scenarios

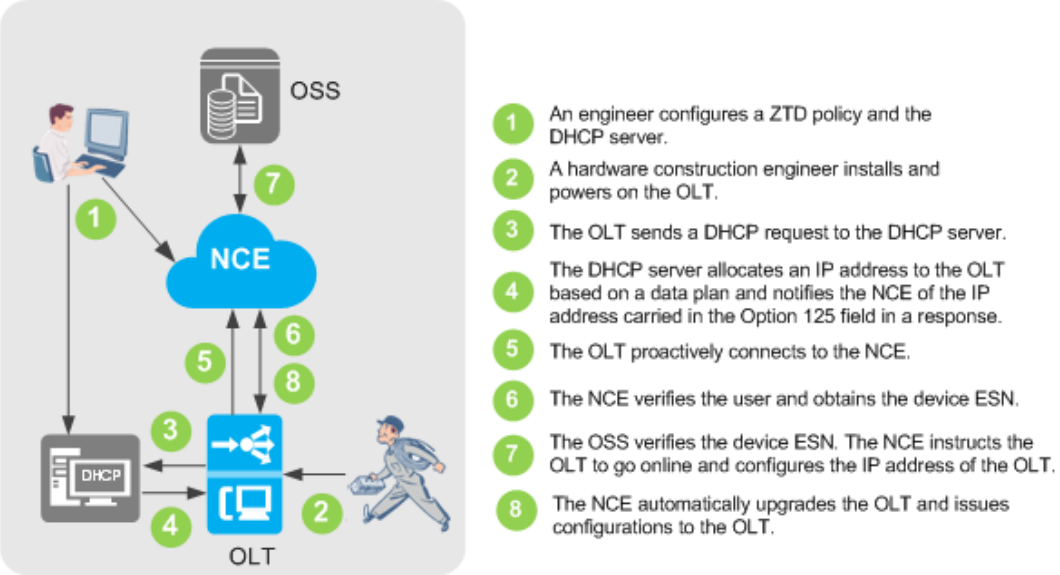
OLT Initial Installation Without the Need of commissioning

After an OLT is powered on and passes cloud-based verification, the NCE automatically upgrades the OLT and issues configurations to the OLT based on a

 Copyright © Huawei Technologies Co., Ltd.

preconfigured process. After configuring basic NE communication parameters on the NCE, an installation and maintenance engineer only needs to install and power on the OLT to provision services and does not need to perform onsite software commissioning.

Figure 19-29 Process of OLT initial installation without the need of commissioning



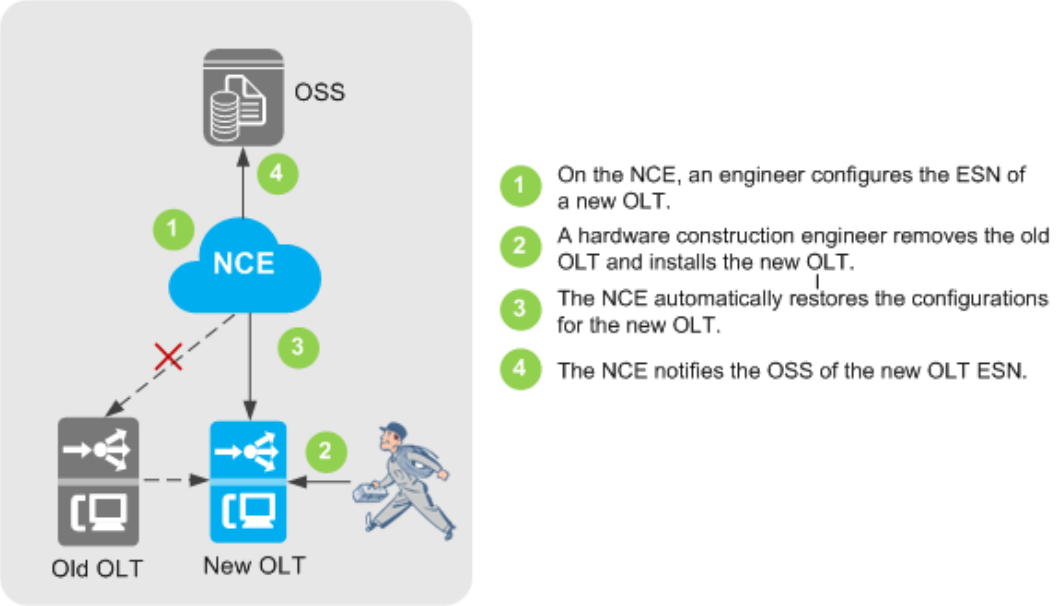
OLT Replacement Due to a Fault

When an OLT needs to be completely replaced because of a subrack fault or when the backplane needs to be replaced, it is required that the software version and data remain the same after the replacement is complete.

On the NCE interface, enter the equipment serial number (ESN) of the old and new OLTs. Replace the correct OLT at the site directly. Then services can be restored through the Call Home power-on process.

 Copyright © Huawei Technologies Co., Ltd.

Figure 19-30 OLT replacement process

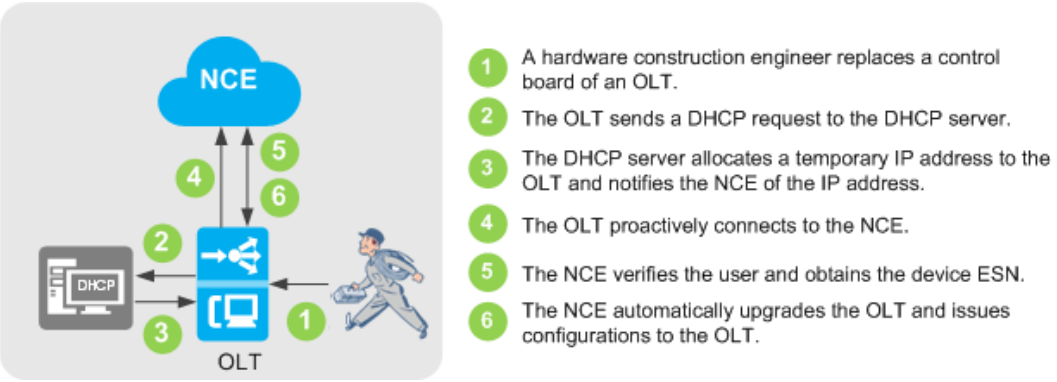


OLT Control Board Replacement Due to a Fault

If control board replacement is required when both the active and standby control boards of an OLT are faulty or when the only control board of an OLT is faulty, the software version and data must remain the same after the replacement is complete.

After a control board is replaced, the NCE applies the original configurations to the OLT based on the same ESN to recover services.

Figure 19-31 OLT Control Board Replacement Process



19.9 Power Saving

 Copyright © Huawei Technologies Co., Ltd.

19.9.1 Introduction to Energy Saving

Definition

The energy saving technology enables the system to automatically detect the usage or service load of each functional module, board, and interface unit of a device. When a module is idle or the load is light, the system automatically disables the module or enables the module to enter low energy consumption mode to reduce the energy consumption of the device.

Purpose

With the surge of bandwidth requirements of broadband users, fiber-in copper-out has accelerated and FTTH, as the current and future mainstream fixed broadband access mode, has a substantial market space. As a result, optical access devices consume significantly more power than other fixed access devices. To reduce such energy consumption and carbon emission, and build a greener ultra-broadband access network, the FTTH energy saving solution is developed.

Benefits

The solution aims to reduce the energy consumption and carbon emission of devices, thereby reducing device use costs.

19.9.2 Energy Saving Principles

19.9.2.1 Energy Saving Principles

To simplify the application and management of the energy saving solution and as different energy saving measures have different impacts on service quality and O&M, an OLT supports five device-level energy saving modes: standard mode, basic mode (default), deep mode, optimal mode, and custom mode. The standard mode supports the fewest energy saving measures but does not have any impact on products. The optimal mode prioritizes energy saving but may affect product service quality and O&M. The custom mode supports flexible configuration of various energy saving measures as required by customers.

Mode	Function Description	Impact on Services
Standard mode	An OLT supports dynamic speed adjustment only for fans. For other hardware components, an OLT supports only static energy saving measures. For example, during device or board initialization, an OLT permanently disables the redundant interfaces and function units of the chips.	This mode does not have any adverse impact on the product. It will not be perceived by users and does not affect user experience.

 Copyright © Huawei Technologies Co., Ltd.

Mode	Function Description	Impact on Services
Basic mode	In addition to the measures supported by the standard mode, the basic mode supports configuration-based dynamic energy saving. For example: • **19.9.2.2.1 Automatic Shutdown of Unused High-speed Interfaces** • **19.9.2.2.2 Automatic Zone-based Fan Speed Adjustment**	This mode does not have any adverse impact on the product. It will not be perceived by users and does not affect user experience.
Deep mode	In addition to the measures supported by the basic mode, the deep mode supports application-based dynamic energy saving. For example: • **19.9.2.2.3 Automatic Shutdown of Unused Ports** • **19.9.2.2.4 Automatic Shutdown of Unused Boards**	This mode has certain impact on the O&M. For example: it will take 2 to 3 minutes to restart a disabled board, and 1 to 2 seconds to enable a disabled port. The electronic label of a disabled board cannot be queried. The mode does not have adverse impact on service quality, but affects new service rollout time due to board and port restart and energy saving configuration check and execution. Port level: The time for new service rollout on a disabled port is prolonged. (by about 1 to 20 seconds). Board level: The time for new service rollout on a disabled board is prolonged (by about 2 to 3 minutes). The electronic label of a disabled board cannot be queried.

 Copyright © Huawei Technologies Co., Ltd.

Mode	Function Description	Impact on Services
Optimal mode	In addition to the measures supported by the deep mode, the optimal mode supports service load-based dynamic energy saving. See **19.9.2.2.6 Dynamic Frequency Adjustment**.	The optimal mode enables a device to run with optimal energy efficiency, but may decrease service performance and quality. For example, in the case of burst traffic, the mode may cause momentary packet loss or bit errors. **NOTE** The mode is not recommended for normal situations and is suitable only for scenarios where energy efficiency is a top priority.
Custom mode	In this mode, in addition to some default energy saving measures (such as automatic fan speed adjustment), users can customize other measures, such as: • **19.9.2.2.3 Automatic Shutdown of Unused Ports** • **19.9.2.2.4 Automatic Shutdown of Unused Boards** • **19.9.2.2.5 Automatic Shutdown of Unused Channels** • **19.9.2.2.7 Intelligent Board Scheduling**	The impact of the mode is determined by customized measures. Disabling unused ports: This measure does not have adverse impact on existing services, but affects the time of new service rollout on disabled ports (by about 1 to 20 seconds). Disabling unused boards: This measure does not have adverse impact on existing services, but affects the time of new service rollout on disabled boards (by about 2 to 3 minutes). The electronic label of a disabled board cannot be queried. Disabling unused channels: This measure does not have adverse impact on existing services, but affects the time of new service rollout in disabled channels (by about 1 to 20 seconds). Intelligent board scheduling: in the case of burst traffic, the mode may slightly increase the latency. **NOTE** The custom mode can be configured only after the corresponding license is applied for on the NMS, and cannot be directly configured on the CLI.

 Copyright © Huawei Technologies Co., Ltd.

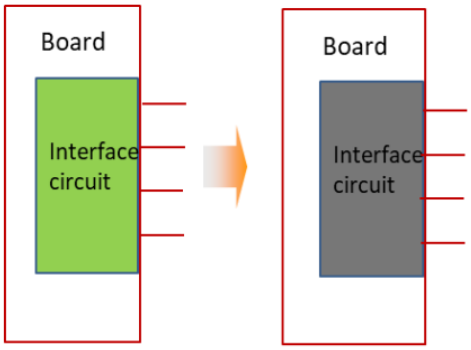
19.9.2.2 Energy Saving Technology Description

19.9.2.2.1 Automatic Shutdown of Unused High-speed Interfaces

An OLT working in basic energy saving mode can detect the usage of high-speed interfaces. When a high-speed interface is not in use, the device automatically shuts down the interface to reduce energy consumption. When the interface needs to be used, the device automatically wakes it up. This does not affect the use of the product. For example:

- When detecting that a service board is not installed in a slot, the system automatically disables the control board high-speed interface that connects to this slot, thereby reducing energy consumption of the control board.
- When detecting that an optical module is not installed on a PON port, the system automatically disconnects the interface circuit of the PON port.

Take a control board for example. The energy saving technology can save 40% energy consumption of a control board in idle state.



Detection scope: all service board slots of a subrack

Detection mechanism:

1. The system checks all slots every several seconds. When detecting that a service board is not installed in a slot or powered off, the system disables the control board high-speed interface that connects to the slot service board.

2. After the board is inserted, the board registration process automatically enables the high-speed interface.

3. During board configuration restoration, if the system detects that an optical module is not installed, it disables the high-speed interface circuit on the board.

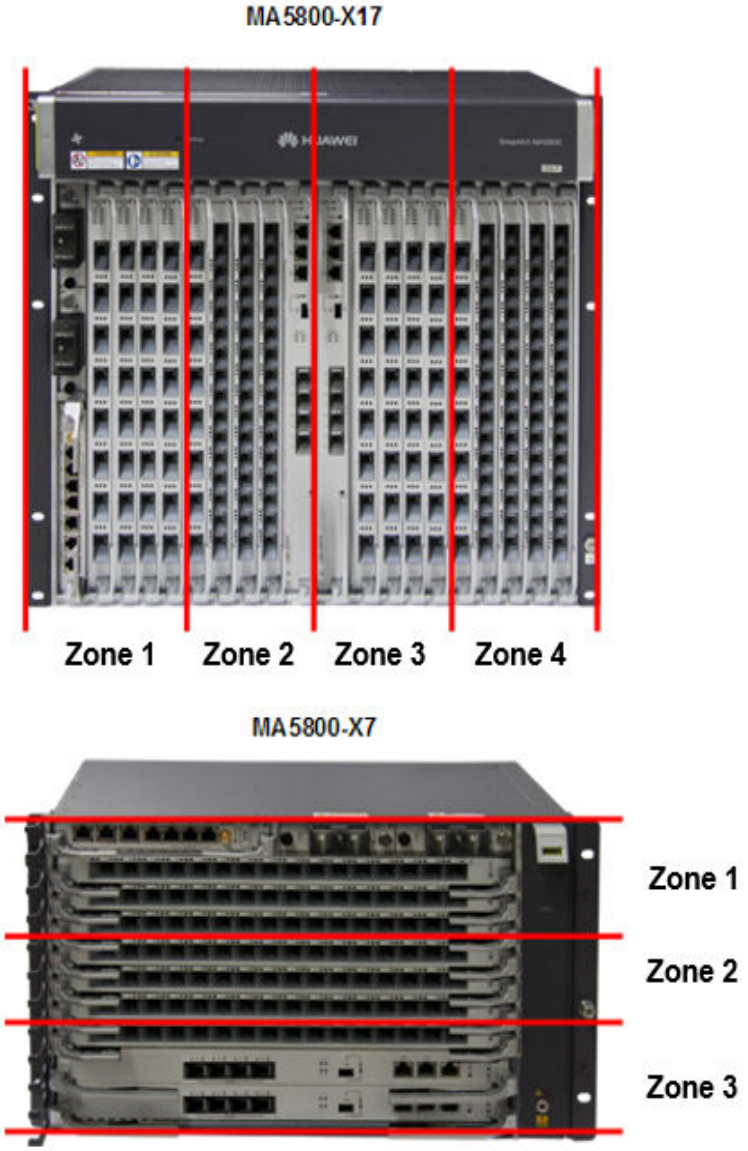
4. After the optical module is removed or inserted, the system detects the installation status change of the optical module and automatically disables or enables the high-speed interface on the board.

19.9.2.2.2 Automatic Zone-based Fan Speed Adjustment

Fans in the subrack of an OLT do not run at a monotonous speed. Instead, an OLT working in basic energy saving mode collects information through the control

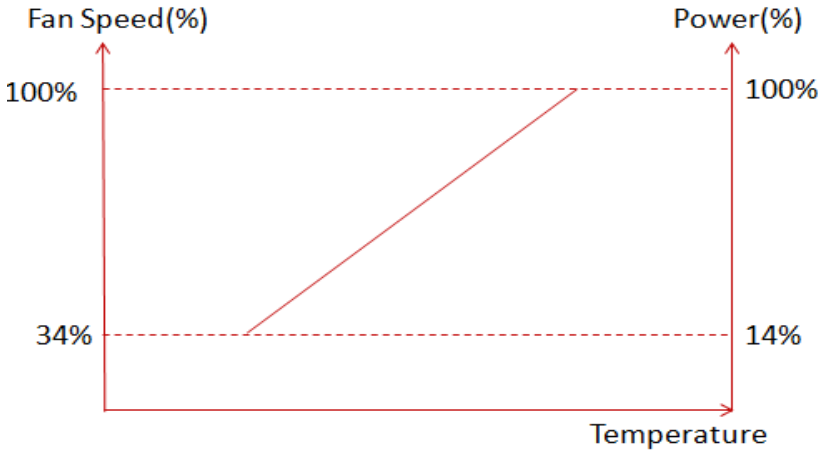
 Copyright © Huawei Technologies Co., Ltd.

board about service board temperatures and zones where the service boards reside. According to the current and target temperatures of each board, the device implements stepless speed adjustment by zone (as shown in the following figure). In this way, fan noises and energy consumption are minimized without affecting reliable running of the device.



The following figure shows the relationships between the fan speed, temperature, and energy consumption.

 Copyright © Huawei Technologies Co., Ltd.



Detection scope: subracks divided into multiple zones

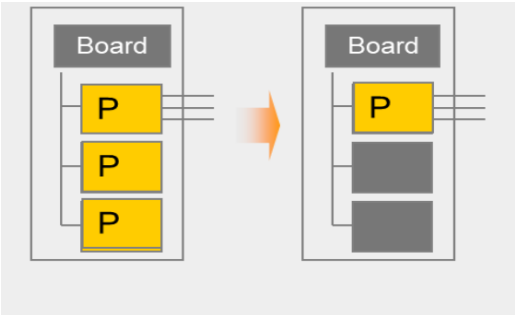
Detection mechanism: The system collects information about temperatures of all normal boards every several seconds and sends it to the fans. Then the fans implement stepless speed adjustment by zone based on the current and target temperatures of each board.

Exceptions:

1. Some subracks only have one zone and do not support this function.

19.9.2.2.3 Automatic Shutdown of Unused Ports

An OLT working in deep energy saving mode can detect the usage of PON board ports. When detecting an unused port, the device automatically shuts down the interface circuits and optical module of the port to reduce energy consumption. When the port needs to be used, the device automatically enables the port and restores it to the normal running state. This does not have any impact on existing service, but may affect new service rollout time on the port (by about 1 to 20 seconds). This technology saves energy consumption by 1.1 W per port for GPON and 1 W per port for Combo PON.



Definition of an unused port:

If ONT auto-discovery check is enabled, a port is considered an unused port when the following conditions are met.

1. No ONT is added to the port.

2. The port does not belong to a Type B or Type C dual-homing protection group.

 Copyright © Huawei Technologies Co., Ltd.

3. The port is not configured with ONT auto-discovery.

If ONT auto-discovery check is disabled (**ont autofind energy-saving enable**), a port is considered an unused port when the following conditions are met:

1. No ONT is added to the port.

2. The port does not belong to a Type B or Type C dual-homing protection group.

Detection scope: GPON or EPON ports only

Detection mechanism: Generally, when users enable automatic shutdown of unused ports or switch to the deep or optimal energy saving mode, the system checks all GPON and EPON ports on each normal board every several seconds. When detecting an unused port, the system disables it. When detecting that a port that needs to be used is shut down for energy saving, the system enables it.

Exceptions:

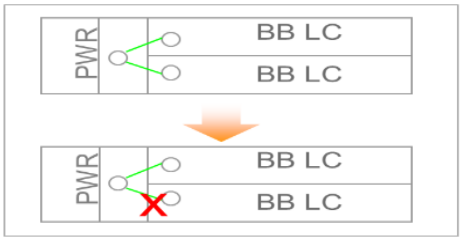
1. If a user disables automatic shutdown for a specific port (**system energy-saving disable f/s/p**), the port will not be automatically disabled for energy saving.

2. If the energy saving mode of a port is switched to manual shutdown (**system energy-saving mode [deep| optimal | custom] method unused-port-shutdown shutdown-operator manual**), the port will not be automatically disabled for energy saving.

3. An L-type board does not support this function.

19.9.2.2.4 Automatic Shutdown of Unused Boards

An OLT working in deep energy saving mode can detect the usage of boards. When detecting an unused board, the OLT automatically shuts down power supply of the board to ensure zero energy consumption. When the board needs to be used, the OLT automatically powers on the board and restores it to the normal running state. This does not have any impact on existing service, but restarting a board may affect the new service rollout time on the board (by about 2 to 3 minutes). This technology can reduce the energy consumption by 100%.



Definition of an unused board:

An ETH access board is considered an unused board when the following conditions are met:

1. The board is not configured with a protection group.

2. There is no service flow on the board.

 Copyright © Huawei Technologies Co., Ltd.

A PON access board is considered an unused board when the following conditions are met:

1. The board is not configured with a protection group.

2. There is no service flow on the board.

3. All ports are unused ports (see the definition of unused port), that is, no ONT is added to any port, no Type B or Type C dual-homing protection group is configured on any port, and ONT auto-discovery is not configured or checked.

Detection scope: PON boards and ETH access boards

Detection mechanism:

The system checks all service boards every several seconds. When detecting that all ports on an idle board without any configuration change in 2 minutes have been disabled for more than 15 minutes, the system shuts down the board. When detecting that the board service configurations change, the system powers on the board immediately.

Exceptions:

1. If a user disables automatic shutdown for a specific board (**system energy-saving disable f/s**), the board will not be automatically disabled for energy saving.

2. If the energy saving mode of a board is switched to manual shutdown (**system energy-saving mode [deep| optimal | custom] method unused-board-shutdown shutdown-operator manual**), the board will not be automatically disabled for energy saving.

3. An L-type board does not support this function.

19.9.2.2.5 Automatic Shutdown of Unused Channels

If automatic shutdown of unused channels is enabled on an OLT working in custom energy saving mode, the OLT detects boards with multiple channels. When detecting an unused channel, the device automatically shuts down the laser of the channel to reduce energy consumption. When detecting that the channel needs to be used, the device automatically enables the channel and restores it to the normal running state. This does not affect the use of the board. This technology can reduce the energy consumption by 0.3 W per channel.

Definition of an unused channel:

If the port where the channel resides is not configured with a protection group and no ONT is added to the channel, the channel is considered an unused channel.

If the port where the channel resides is configured with a protection group and no ONT is added to the channel of the master port, the channel is considered an unused channel.

Detection scope: boards with multiple channels

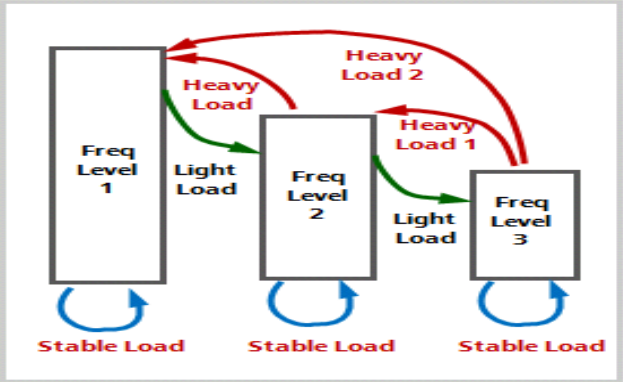
Detection mechanism: When a user switches to the custom mode and enables automatic shutdown of unused channels, the system checks all channels of each

 Copyright © Huawei Technologies Co., Ltd.

normal board every several seconds. When detecting an unused channel, the system shuts it down. When detecting that a channel that needs to be used is shut down for energy saving, the system enables it.

19.9.2.2.6 Dynamic Frequency Adjustment

An OLT working in optimal energy saving mode can detect the service load of some service boards. When detecting that a service board is carrying a small service load, the device automatically reduces the working frequency of the service processing module of the board to reduce energy consumption. When the service load of the board increases, the device automatically increases the working frequency of the service processing module to meet service needs. This technology can reduce the energy consumption by 2 to 3 W per board.



CAUTION

In the case of burst traffic, this function may cause momentary packet loss or bit errors.

19.9.2.2.7 Intelligent Board Scheduling

An OLT working in custom energy saving mode can detect the service load of some service boards. When detecting that a service board is carrying a small service load, the device automatically reduces the working frequency of the service processing module of the board to reduce energy consumption. When the service load of the board increases, the device automatically increases the working frequency of the service processing module to meet service needs.

This technology can reduce the energy consumption by 2 to 3 W per board.

CAUTION

In the case of burst traffic, this function will not cause packet loss but may increase the latency for a short time.

 Copyright © Huawei Technologies Co., Ltd.

19.9.3 Energy Saving Modes and Constraints

- The following technologies can be used to configure the custom energy-saving mode: automatic shutdown of unused ports, automatic shutdown of unused boards, intelligent board scheduling, and automatic shutdown of unused channels.
- Automatic shutdown of unused high-speed interfaces and automatic fan speed adjustment by zone are enabled in custom mode and cannot be disabled.
- The dynamic frequency adjustment technology may cause packet loss in the case of burst traffic.
- After an optical module is removed and re-inserted, the board does not report an alarm indicating that the optical module is not in position if board energy saving is enabled.
- If an alarm indicating that the optical module is not in position needs to be reported after port energy saving is enabled, you need to run the **xpon alarm optical-module-alarm control disable** command.

19.9.4 Principles of Power Savings and Power Estimation

1. Principles for Estimating OLT Power Savings

   Estimated power savings = Power saved by shutting down boards + Power saved by shutting down ports + Power saved by shutting down channels

   - **Shutdown of unused boards**: Calculate the difference between the power before and after a board is shut down. For example, if the total power is 300 W before a board is shut down and 250 W after the board is shut down, the power saved by shutting down the board is 50 W (with a power fluctuation error).
   - **Shutdown of unused ports**: Based on different combinations of the board type, port mode, and optical module in-position status, the typical value tested in the lab is used for estimation. For example, for the H908CGHF board, if the port mode is XG/G PON and the optical module is in position, shutting down one port saves 1.1 W (lab test data) and shutting down five ports saves 5.5 W.
   - **Shutdown of unused channels:** Based on different combinations of the board type, port mode, channel type, and optical module in-position status, the typical value tested in the lab is used for estimation. For example, for the H908CGHF board, if port mode is XG/G PON and the optical module is in position, shutting down one 10G PON channel saves 0.8 W (lab test data) and shutting down one GPON channel saves 0.2 W (lab test data).

2. Description of OLT power saving errors

   - **The error in power saved by shutting down unused boards comes from the current collection precision of the power board.**

     The power saved for the entire subrack is calculated by subtraction, which depends on the precision of the current collected by the power board. The collection precision meets the ETSI 202 336-12 standard.

     Power board load at 25%–100%: ±3%

     Power board load at 5%–25%: ±5%

 Copyright © Huawei Technologies Co., Ltd.

- **The error in power saved by shutting down unused ports comes from optical modules.**

  Data source of the typical power of optical modules: The data of each type of optical modules is the average value of 16 optical modules of the same type selected from mainstream vendors and versions in the lab under the normal temperature of 25°C.

  The power of an optical module is affected by the following factors:

  i. **Tx optical power:** The power varies with the optical power of optical modules (Class B+, Class C+, and Class D).

  ii. **Process**: The power of optical modules varies with the process control (such as the processing precision, temperature, and humidity of the production line) of different vendors and batches.

  iii. **Temperature**: The power of an optical module varies with the temperature. Generally, the higher the temperature, the larger the leakage current and the higher the power.

19.9.5 Energy Saving Configuration Description

19.9.5.1 Energy Saving Configuration

There are to ways to configure energy saving. One is to configure the 4 energy saving modes (standard, basic, deep, and optimal modes) on an OLT, and the other is to configure the custom mode and specific energy saving measures on the network management system (NMS) through the network cloud engine (NCE).

19.9.5.1.1 Configuring an Energy Saving Mode

Run the **system energy-saving mode** command to set the energy saving mode of the system.

Run the **undo system energy-saving mode** command to restore the energy saving mode of the system to the default value (basic mode).

Run the **display system energy-saving mode** command to query the current energy saving mode of the system.

To set the current energy saving mode of the system to deep mode, do as follows:

```
huawei(config)#system energy-saving mode deep
{ <cr>|method<K> }:
  Command:
        system energy-saving mode deep
  Command is being executed, please wait for a moment...
  Set the energy-saving mode successfully
```

Table 19-15 The following table lists the energy saving measures corresponding to each mode.

Energy Saving Mode	Measures
Standard mode	19.9.2.2.1 Automatic Shutdown of Unused High-speed Interfaces

 Copyright © Huawei Technologies Co., Ltd.

Energy Saving Mode	Measures
Basic mode	• 19.9.2.2.1 Automatic Shutdown of Unused High-speed Interfaces • 19.9.2.2.2 Automatic Zone-based Fan Speed Adjustment
Deep mode	• 19.9.2.2.1 Automatic Shutdown of Unused High-speed Interfaces • 19.9.2.2.2 Automatic Zone-based Fan Speed Adjustment • 19.9.2.2.3 Automatic Shutdown of Unused Ports • 19.9.2.2.4 Automatic Shutdown of Unused Boards
Optimal mode	• 19.9.2.2.1 Automatic Shutdown of Unused High-speed Interfaces • 19.9.2.2.2 Automatic Zone-based Fan Speed Adjustment • 19.9.2.2.3 Automatic Shutdown of Unused Ports • 19.9.2.2.4 Automatic Shutdown of Unused Boards • 19.9.2.2.6 Dynamic Frequency Adjustment
Custom mode	Measures that take effect by default: • 19.9.2.2.1 Automatic Shutdown of Unused High-speed Interfaces • 19.9.2.2.2 Automatic Zone-based Fan Speed Adjustment Measures that are enabled manually: • 19.9.2.2.3 Automatic Shutdown of Unused Ports • 19.9.2.2.4 Automatic Shutdown of Unused Boards • 19.9.2.2.5 Automatic Shutdown of Unused Channels • 19.9.2.2.7 Intelligent Board Scheduling **CAUTION** You need to apply for a license and then configure the energy saving mode on the NCE. It cannot be directly configured on the CLI.

Copyright © Huawei Technologies Co., Ltd.

19.9.5.1.2 Configuring Custom Energy Saving Measures

Users can run the **system energy-saving mode custom method** method-name { **enable** | **disable** } command to enable or disable energy saving measures in custom mode.

Table 19-16 Description of the *method-name* parameters

Parameter	Measure Name
unused-board-shutdown	Indicates that a board is shut down when not in use to save energy.
unused-port-shutdown	Indicates that a port is shut down when not in use to save energy.
intelligent-board-scheduling	Indicates intelligent board scheduling to save energy.
unused-channel-shutdown	Indicates that a channel is shut down when not in use to save energy.

Configuring Shutdown of Unused Boards

To enable shutdown of unused boards in custom energy saving mode, do as follows:

```
huawei(config)#system energy-saving mode custom method unused-board-shutdown enable
  Set the energy-saving method successfully
```

Configuring Shutdown of Unused Ports

To enable shutdown of unused ports in custom energy saving mode, do as follows:

```
huawei(config)#system energy-saving mode custom method unused-port-shutdown enable
  Set the energy-saving method successfully
```

Configuring Parameters for Unused Port Shutdown: Disabling ONT Auto-Discovery Check

Generally, ONT auto-discovery is regarded as a port configuration. After ONT auto-discovery is configured (**port ont-auto-find**) on a port, the port cannot be shut down for energy saving even though it is not configured with any other services. To enable both ONT auto-discovery and port shutdown when not in use, run the **ont autofind energy-saving enable** command to disable ONT auto-discovery check.

You can run the **display ont global-config** or **display xpon global-config** command to query whether ONT auto-discovery check is enabled.

To enable unused port shutdown in custom energy saving mode and disable ONT auto-discovery check, do as follows:

```
huawei(config)#system energy-saving mode custom method unused-port-shutdown enable
  Set the energy-saving method successfully

huawei(config)#ont autofind energy-saving enable
```

 Copyright © Huawei Technologies Co., Ltd.

To disable ONT auto-discovery check in deep energy saving mode, do as follows:

```
huawei(config)#system energy-saving mode deep
{ <cr>|method<K> }:

 Command:
       system energy-saving mode deep
 Command is being executed, please wait for a moment...
 Set the energy-saving mode successfully

huawei(config)#ont autofind energy-saving enable
```

Enable deep energy saving for an optical module. This function is supported only in V100R022C10SPH211 version.

```
huawei(config)#xpon global-config switch optical-deep-energy-saving enable
```

Configuring Shutdown of Unused Channels

To enable unused channel shutdown in custom energy saving mode, do as follows:

```
huawei(config)#system energy-saving mode custom method unused-channel-shutdown enable
 Set the energy-saving method successfully
```

Configuring Intelligent Board Scheduling

To enable intelligent board scheduling in custom energy saving mode, do as follows:

```
huawei(config)#system energy-saving mode custom method intelligent-board-scheduling enable
 Set the energy-saving method successfully
```

19.9.5.1.3 Querying Energy Saving Configurations

1. Run the **display system energy-saving mode** command to query the current energy saving mode.

```
huawei#display system energy-saving mode
{ <cr>||<K> }:
 Command:
       display system energy-saving mode
 Energy-saving mode: basic
```

2. Run the **display system energy-saving method custom** command to query energy saving measures in custom mode.

```
huawei#display system energy-saving method custom
{ <cr>||<K> }:
 Command:
 display system energy-saving method custom
 ---------------------------------------------------------------------------
 Method                              Custom
 ---------------------------------------------------------------------------
 zone-based-fan-speed-adjustment        Y
 unused-high-speed-interface-shutdown   Y
 unused-board-shutdown                  N
 unused-port-shutdown                   N
 unused-channel-shutdown                N
 intelligent-board-scheduling           N
 ---------------------------------------------------------------------------
```

3. Run the **display system energy-saving parameter** command to query parameters of a specified measure.

```
huawei#display system energy-saving parameter
{ <cr>||<K> }:
```

 Copyright © Huawei Technologies Co., Ltd.

```
  Command:
       display system energy-saving parameter
  ------------------------------------------------------------
  Mode:Deep
  ------------------------------------------------------------
  Method                    Parameter          Value
  ------------------------------------------------------------
  unused-board-shutdown      shutdown-operator   auto
  unused-port-shutdown       shutdown-operator   auto
  ------------------------------------------------------------
  ------------------------------------------------------------
  Mode:Optimal
  ------------------------------------------------------------
  Method                    Parameter          Value
  ------------------------------------------------------------
  unused-board-shutdown      shutdown-operator   auto
  unused-port-shutdown       shutdown-operator   auto
  ------------------------------------------------------------
  ------------------------------------------------------------
  Mode:Custom
  ------------------------------------------------------------
  Method                    Parameter          Value
  ------------------------------------------------------------
  unused-board-shutdown      shutdown-operator   auto
  unused-port-shutdown       shutdown-operator   auto
  ------------------------------------------------------------
```

Run the **display ont global-config** or **display xpon global-config** command to query whether ONT auto-discovery check is enabled.

```
huawei#display ont global-config
{ <cr>||<K> }:
  Command:
       display ont global-config
  ------------------------------------------------------------------------------
  ONT autofind aging time(s)                    : 300
  ONT autofind conflict check switch             : disable
  ONT autofind distance time(only for GPON)(s)    : 5
  ONT autofind energy-saving                    : enable
  ONT CATV port current default operational state : on
  ONT auto-migration switch                     : disable
  ONT auto-aging time(day)                      : no-aging
  ONT authentication conflict check by-sn        : enable
  ONT authentication conflict check by-loid      : enable
  ONT autofind additional notice switch          : disable
  ONT autofind additional notice times           : -
  ONT autofind additional notice interval(s)     : -
  ONT online rate control                       : disable
  ------------------------------------------------------------------------------
huawei#display xpon global-config
{ <cr>||<K> }:

  Command:
       display xpon global-config
  ------------------------------------------------------------------------------
  Current configuration mode                    : distributing-mode
  ------------------------------------------------------------------------------
  xPON rogue ONT auto detection switch           : off
  xPON alarm clear-on-shutdown switch            : disable
  xPON group power off alarm switch              : disable
  xPON LOS alarm control switch                 : disable
  xPON LOS alarm control DGi threshold ratio(%)   : 20
  xPON LOS alarm control DGi threshold time(s)    : 10
  xPON suppress initial alarm switch             : enable
  xPON suppress optical initial alarm switch     : disable
  xPON optical module alarm control switch       : disable
  xPON distribute fiber LOS alarm switch         : disable
  xPON modify bound profile switch               : enable
  xPON protect group different range switch      : disable
  xPON port color sensitive switch              : disable
```

 Copyright © Huawei Technologies Co., Ltd.

```
xPON ONT power reduction switch                : enable
xPON ONT offline monitor switch                : disable
xPON ONT offline monitor alarm threshold(%)    : 70
xPON ONT offline monitor restore threshold(%)  : 80
xPON auto-alloc GEM port encrypt               : off
xPON virtual system ONT ID alloc policy        : independent
xPON optical-alarm auto-monitor port switch    : standard-threshold
xPON optical-alarm auto-monitor ont switch     : disable
xPON ONT IPv6 multicast switch in CTC          : disable
xPON ONT optical-ddm quick-query switch        : disable
xPON ONT distance variation threshold(m)       : Unlimited
xPON rogue ONT detect-rssi-threshold(dBm)      : -28.00
xPON rogue ONT clear-rssi-threshold(dBm)       : -28.00
xPON bandwidth combine switch                  : disable
xPON default OMCC encrypt switch               : enable
xPON default GEM port encrypt switch           : enable
xPON OLT RX ONT power mode                     : high-precision
xPON ONT type check strict switch              : disable
xPON channel rogue onu check switch            : enable
xPON optical-module info mode                  : normal
xPON SIP URI format                            : TEL
xPON operator check CLI switch                 : enable
xPON mode2 time sync switch                    : disable
xPON optical module deep-energy-saving switch  : enable
------------------------------------------------------------------------------
ONT autofind aging time(s)                     : 300
ONT autofind conflict check switch             : disable
ONT autofind distance time(only for GPON)(s)   : 5
ONT autofind energy-saving                     : enable
ONT CATV port current default operational state : on
ONT auto-migration switch                      : enable
ONT auto-aging time(day)                       : 15
ONT authentication conflict check by-sn        : enable
ONT authentication conflict check by-loid      : enable
ONT autofind additional notice switch          : enable
ONT autofind additional notice times           : 5
ONT autofind additional notice interval(s)     : 100
ONT online rate control                        : disable
------------------------------------------------------------------------------
GPON DBA bandwidth type                        : GPON
GPON DBA bandwidth assignment mode             : min-loop-delay
GPON multicast GEM port                        : 4095
XG-PON multicast GEM port                      : 65532
GPON ONT password renew interval(min)          : -
GPON T-CONT auto-create switch                 : disable
GPON interoperability standard                 : ITU-T
GPON interoperability multicast authentication : ONT control
GPON interoperability T-CONT PQ priority reverse: disable
GPON interoperability ONT WAN config mode      : set-by-create
GPON ONT home gateway config method            : default
GPON anti-dos PLOAM switch                     : enable
GPON ONT SN format                             : vendor-prefix
GPON ONT vlan tag operation mode               : modify single tag
GPON ONT multicast vlan filter                 : default
XGS-PON RTU allocation priority policy         : Bandwidth RTU first
GPON broadcast GEM port                        : disable
Extended VLAN tagging operation                : enable
ONT ETH layer-2 service switch                 : enable
------------------------------------------------------------------------------
```

NOTE

```
xPON SIP URI format                            : TEL
xPON operator check CLI switch                 : enable
xPON mode2 time sync switch                    : disable
xPON optical module deep-energy-saving switch  : enable
```

Supported only in V100R022C10SPH211 version.

 Copyright © Huawei Technologies Co., Ltd.

19.9.5.1.4 Powering Off an Unused Port/Board

The **system energy-saving power-off** { frameid/slotid / frameid/slotid/portid } command is used to power off a board or port. When you need to shut down a board or port that does not carry any services to save energy, run this command.

The **system energy-saving power-on** { frameid/slotid / frameid/slotid/portid } command is used to power on a board or port that has been powered off for energy saving. When you need to configure services for a board or port that has been shut down for energy saving, run this command.

To power off board 0/1 for energy saving, do as follows:

```
huawei(diagnose)%%system energy-saving power-off 0/1
  Are you sure to power off the board?(y/n)[n]:y
  Energy saving shutdown successfully
```

19.9.5.2 Querying Energy Saving Results

19.9.5.2.1 Querying Energy Saving Status

The **display system energy-saving info** { frameid | frameid/slotid } command is used to query the energy saving running status of a specified subrack or board.

The **display system energy-saving info detail** { frameid/slotid | frameid/slotid/portid } command is used to query detailed energy saving information about a specified board or port.

Parameter	Description
Frame/Slot	Indicates the subrack ID and slot ID of a board.
Frame/Slot/Port	Indicates the subrack ID, slot ID, and port ID of a port.
Enabled	Indicates the enabling status of the energy-saving function of a board or port. You can run the **system energy-saving** command to set this parameter.
EnergySavingOperStatus	Indicates the energy saving running status of a board or port. Options: • **NoEnergySaving**: no energy saving state. • **Deactived**: deactivated state. • **PartialEnergySaving**: partial energy saving state. • **ESavingStandby**: shutdown when no service is running. When a board or port is in this state, you can run the **system energy-saving power-off** command to configure shutdown measures for energy saving. • **ESavingShutdown**: energy saving shutdown state. You can run the **system energy-saving power-off** command to set this parameter. • **ManualShutdown**: manual shutdown state. You can run the **board power-off** command to set this parameter.

 Copyright © Huawei Technologies Co., Ltd.

Parameter	Description
Channel x EnergySavingOperStatus	Indicates the energy saving running status of a channel. Options: • **NoEnergySaving**: no energy saving state. • **ESavingShutdown**: energy saving shutdown state. • **ManualShutdown**: manual shutdown state. You can run the **port laser-select(gpon)** or **port laser-select(epon)** command to set this parameter. This information is displayed only when the detailed information about the energy-saving function of a port is queried on a board that supports dual channels. Replace *x* with 0 or 1.

19.9.5.2.2 Querying the Shutdown Status of a Board

The **display system energy-saving info** { frameid | frameid/slotid } command is used to query the energy saving running status of a specified board or port.

To query the brief information about the energy-saving function of all boards in subrack 0, do as follows:

```
huawei#display systeEm energy-saving info 0
{ <cr>||<K> }:
  Command:
       display system energy-saving info 0
  ------------------------------------------------------------
  Frame/Slot         Enabled   EnergySavingOperStatus
  ------------------------------------------------------------
  0/0              Enable    NoEnergySaving
  0/1              Enable    NoEnergySaving
  0/2              Enable    NoEnergySaving
  0/3              Enable    ESavingShutdown
  0/4              Enable    ESavingShutdown
  0/5              Enable    NoEnergySaving
  0/9              Enable    NoEnergySaving
  0/10              Enable    NoEnergySaving
  ------------------------------------------------------------
```

As shown in the command output, subracks 0/3 and 0/4 are shut down for energy saving.

19.9.5.2.3 Querying the Shutdown Status of a Port

To query the brief information about the energy-saving function of all ports on board 0/1, do as follows:

```
huawei#display system energy-saving info 0/1
{ <cr>||<K> }:
  Command:
       display system energy-saving info 0/1
  ------------------------------------------------------------
  Frame/Slot/Port     Enabled   EnergySavingOperStatus
  ------------------------------------------------------------
  0/1/0            Enable    NoEnergySaving
  0/1/1            Enable    NoEnergySaving
  0/1/2            Enable    ESavingShutdown
  0/1/3            Enable    ESavingShutdown
```

 Copyright © Huawei Technologies Co., Ltd.

```
0/1/4           Enable   NoEnergySaving
0/1/5           Enable   NoEnergySaving
------------------------------------------------------------
```

As shown in the command output, ports 0/1/2 and 0/1/3 are shut down for energy saving.

19.9.5.2.4 Querying the Shutdown Status of a Channel

To query the detailed information about the energy-saving function of port 0/1/0, do as follows:

```
huawei#display system energy-saving info detail 0/1/0
{ <cr>||<K> }:
 Command:
      display system energy-saving info detail 0/1/0
 ------------------------------------------------------------
 Enabled: Enable
 EnergySavingOperStatus: PartialEnergySaving
 Channel 0 EnergySavingOperStatus: ESavingShutdown
 Channel 1 EnergySavingOperStatus: ESavingShutdown
 ------------------------------------------------------------
```

As shown in the command output, port 0/1/0 on the board is shut down for energy saving.

19.9.5.2.5 Querying Zone-based Fan Speed Adjustment Results

Run the **display fan speed** command to query the speed of each fan. When board temperature differences are large between zones, the fan speeds change.

To query the current fan speeds, do as follows:

```
huawei(config-if-fan-1)%%display fan speed
Fan 0 Real-time speed: 1470times/minute, Fan speed Occupy:  32%
Fan 1 Real-time speed: 1530times/minute, Fan speed Occupy:  34%
Fan 2 Real-time speed: 1530times/minute, Fan speed Occupy:  34%
Fan 3 Real-time speed: 1530times/minute, Fan speed Occupy:  34%
Fan 4 Real-time speed: 1500times/minute, Fan speed Occupy:  33%
Fan 5 Real-time speed: 1500times/minute, Fan speed Occupy:  33%
Fan 6 Real-time speed: 1500times/minute, Fan speed Occupy:  33%
Fan 7 Real-time speed: 1500times/minute, Fan speed Occupy:  33%
```

As shown in the command output, when board temperature differences are large between zones, the fan speeds change.

19.9.5.2.6 Querying Dynamic Frequency Adjustment Results

For a board that supports dynamic frequency adjustment, if the board version is V100R023C1xSPC200 or earlier, you can only run the expert maintenance command to enter the board CLI through the transparent channel and then run the **asic freq get** command to query the adjustment status.

For example:

In privilege mode or global config mode, run the **diagnose** command to enter diagnose mode.

In diagnose mode, run the **su** command to enter SU mode.

Run the **transparent on 0/1** command to enter the transparent channel of slot 1, and then run the **asic freq get** command.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(su)%%asic freq get
atom: 850
rp  : 425
eqm : 425
bac : 533.33
ifb : 300
pa  : 533.33
pe  : 533.33
pba : 400
```

You can obtain the dynamic frequency adjustment result based on the *atom* frequency change in the command output.

The following table describes the parameters in response to this command:

Parameter	Description
atom	Indicates the ATOM clock frequency.
rp	Indicates the RP clock frequency.
eqm	Indicates the EQM clock frequency.
bac	Indicates the BAC clock frequency.
ifb	Indicates the IFB clock frequency.
pa	Indicates the PA clock frequency.
pe	Indicates the PE clock frequency.
pba	Indicates the PBA clock frequency.

In addition, when dynamic frequency adjustment takes effect for a board, the board is in the partial energy saving state. You can run the **display system energy-saving info detail 0/1** command to query the information.

```
huawei#display system energy-saving info detail 0/1
<cr>||<K> }:
 Command:
       display system energy-saving info detail 0/1
 ------------------------------------------------------------
 Enabled: Enable
 EnergySavingOperStatus: PartialEnergySaving
 ------------------------------------------------------------
```

As shown in the command output, when dynamic frequency adjustment takes effect, the board in subrack 0/1 is in the partial energy saving state.

19.9.5.2.7 Querying Intelligent Board Scheduling Results

For a board that supports intelligent board scheduling, if the board version is V100R023C1xSPC200 or earlier, you can only run the expert maintenance command to enter the board CLI through the transparent channel and then run the **asic power display np-dynamic-freq-info** command to query the intelligent board scheduling status.

For example:

In privilege mode or global config mode, run the **diagnose** command to enter diagnose mode.

 Copyright © Huawei Technologies Co., Ltd.

In diagnose mode, run the **su** command to enter SU mode.

Run the **transparent on 0/2** command to enter the transparent channel of slot 2, and then run the **asic power display np-dynamic-freq-info** command.

```
huawei(su)%%asic power display
{ freq<K>|pll-power-down-en<K>|sds-power-down-en<K>|np-dynamic-freq-info<K>|tm-
dynamic-freq-info<K> }:np-dynamic-freq-info
 en      up-th  down-th freq
 disable 950M   425M    950M
 Command execute success!
```

You can obtain the intelligent board scheduling result based on the *freq* frequency change in the command output.

The following table describes the parameters in response to this command:

Parameter	Description
en	Indicates the NP dynamic frequency adjustment status. • enable • disable
up-th	Indicates the highest frequency.
down-th	Indicates the lowest frequency.
freq	Indicates the current frequency.

In addition, when intelligent board scheduling takes effect for a board, the board is in the partial energy saving state. You can run the **display system energy-saving info detail 0/2** command to query the information.

```
huawei#display system energy-saving info detail 0/2
<cr>||<K> }:
  Command:
        display system energy-saving info detail 0/2
  ------------------------------------------------------------
  Enabled: Enable
  EnergySavingOperStatus: PartialEnergySaving
  ------------------------------------------------------------
```

As shown in the command output, when intelligent board scheduling takes effect, the board in subrack 0/2 is in the partial energy saving state.

19.9.5.3 Energy Saving Configuration Examples

19.9.5.3.1 Configuring the Deep Energy Saving Mode on an OLT

Step 1 Set the current energy saving mode of the system to deep mode.

```
huawei(config)#system energy-saving mode deep
{ <cr>|method<K> }:
  Command:
        system energy-saving mode deep
  Command is being executed, please wait for a moment...
  Set the energy-saving mode successfully
```

Step 2 Check whether the current energy-saving mode takes effect.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei#display system energy-saving mode
{ <cr>||<K> }:
  Command:
        display system energy-saving mode
  Energy-saving mode: deep
```

Step 3 If there are unused boards or ports in the environment, they will be detected after a certain period of time and shut down for energy saving. You can run the **display system energy-saving info** command to query the energy saving result.

```
huawei#display system energy-saving info 0/1
{ <cr>||<K> }:
  Command:
        display system energy-saving info 0/1
  ------------------------------------------------------------
  Frame/Slot/Port     Enabled   EnergySavingOperStatus
  ------------------------------------------------------------
  0/1/0               Enable    ESavingShutdown
  0/1/1               Enable    ESavingShutdown
  0/1/2               Enable    ESavingShutdown
  0/1/3               Enable    ESavingShutdown
  0/1/4               Enable    ESavingShutdown
  0/1/5               Enable    ESavingShutdown
------------------------------------------------------------
```

Step 4 Run the **undo system energy-saving mode** command to exit the deep energy saving mode.

```
Huawei(config)#undo system energy-saving mode
{ <cr>||<K> }:
  Command:
        undo system energy-saving mode
  Command is being executed, please wait for a moment...
  Set the energy-saving mode successfully
```

----End

19.9.5.3.2 Configuring the Deep Energy Saving Mode on the NMS

NMS Version Requirements

- For NEs of V100R022C00 and earlier versions, ONU auto-discovery check disabling is not supported.
- For NEs of V100R022C10SPC200 and earlier versions:

Only the energy saving measures automatic shutdown of unused boards and automatic shutdown of unused ports are supported. The association relationships between the measures are as follows:

When either of automatic shutdown of unused boards and automatic shutdown of unused ports is enabled, the other measure is automatically enabled on same device.

A union set of effective periods of associated energy saving measures is used. As long as one measure is within its energy saving period, the others take effect as well.

- For NEs of V100R022C10SPH203 and later versions, all energy saving functions are supported and there are no associations between them.

 Copyright © Huawei Technologies Co., Ltd.

Procedure

Step 1 Open the Access Controller app and choose **Automation Control Center** from the main menu.

Step 2 In the **Automation Control Center** area, click **Energy Consumption Management**.

Step 3 Click the **Energy Saving Model Management** tab.

Step 4 Use either of the following methods to create an energy saving model.

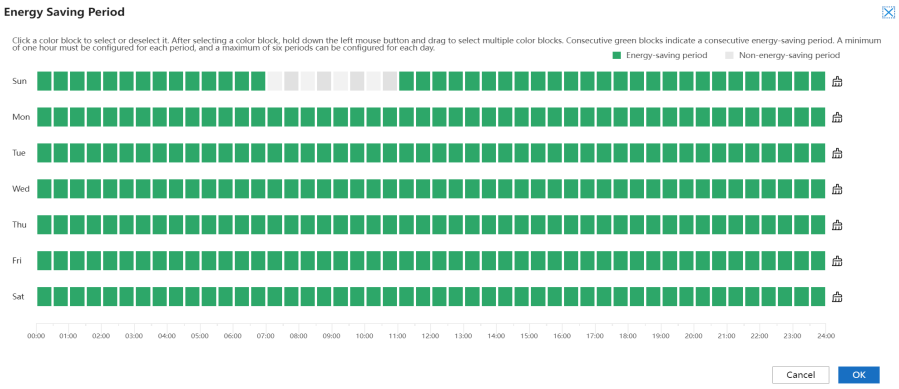
- Configuring an Energy Saving Model
  a. Click **Create**. In the displayed dialog box, set the energy-saving model name, and enable or disable energy-saving measures and ONU auto-discovery check.

Parameter	Description
Automatic shutdown of unused boards	When a board is not in use, it is automatically powered off. When the system detects that the board needs to be used, it automatically powers on the board and restores it to the normal running state. **NOTE** This does not affect service quality but has certain impact on the O&M. It takes about 2 to 3 minutes to restart a board that is shut down.
Automatic shutdown of unused ports	When a PON port is not in use, the corresponding interface circuit and optical module are automatically shut down. When the system detects that the port needs to be used, it automatically enables the port and restores it to the normal running state. **NOTE** This does not affect service quality but has certain impact on the O&M. It takes about 1 to 2 seconds to restart a port that is shut down.
Automatic shutdown of unused channels	When there is an unused channel of a Combo PON port, the channel is automatically shut down. When the system detects that the channel needs to be used, it automatically enables the channel and restores it to the normal running state. **NOTE** This does not affect service quality but has certain impact on the O&M. It takes about 1 to 2 seconds to restart a channel that is shut down.

 Copyright © Huawei Technologies Co., Ltd.

Parameter	Description
Intelligent board scheduling	A board intelligently schedules internal forwarding resources based on the service load. When the service load decreases, a board automatically disables some forwarding resources or reduces the working frequency of forwarding resources. When the service load increases, the board enables the forwarding resources that have been disabled or improves the working frequency of forwarding resources. **NOTE** In the case of burst traffic, packet forwarding may have latency, but no packets will be lost.
ONU auto-discovery check disabling	After this function is enabled, the system does not check the ONU auto-discovery configuration for energy saving measures, that is, automatic shutdown of unused boards, ports, and channels, to take effect. The ONU auto-discovery function is not available when energy saving measures are effective.

b. Click corresponding to an energy saving measure to set the energy saving period.



c. Click **OK**.

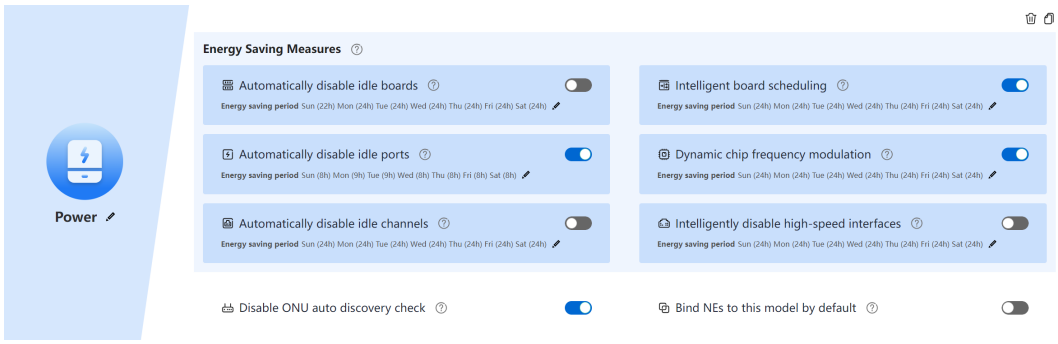
- Copying an Energy Saving Model

  a. Click corresponding to an energy saving model. In the displayed dialogue box, set the energy saving model name, enable or disable energy saving measures and ONU auto-discovery check, and set the energy saving period.

  b. Click **OK**.

Step 5 After an energy saving model is created, you can modify the model name, energy saving measure, and energy saving period. To enable an NE to be automatically bound to the energy saving model when it goes online, enable **Bind NEs to this model by default**.

 Copyright © Huawei Technologies Co., Ltd.



----End

19.9.6 Energy Saving Conditions

19.9.6.1 Function Conditions

19.9.6.1.1 Energy Saving Mode Conditions

Energy Saving Mode	Supported Versions (Replace *x* in the Version Number with 0, 1, or 2.)	NMS Start Version
Standard mode	V100R018C00SPC500, V100R019C1xSPC200, and respective later versions	V100R022C10SPC101
Basic mode	V100R018C00SPC500, V100R019C1xSPC200, and respective later versions	V100R022C10SPC101
Deep mode	V100R018C00SPC500, V100R019C1xSPC200, and respective later versions	V100R022C10SPC101
Optimal mode	V100R018C00SPC500, V100R019C1xSPC200, and respective later versions	V100R022C10SPC101
Custom mode	V100R022C1xSPH203and later versions	V100R022C10SPC101

 Copyright © Huawei Technologies Co., Ltd.

19.9.6.1.2 Energy Saving Measure Conditions

Energy Saving Mode	Energy Saving Measure	Supported Versions (Replace *x* in the Version Number with 0, 1, or 2.)
Standard mode	Zone-based automatic fan speed adjustment	V100R018C00SPC500, V100R019C1xSPC200, and respective later versions
Basic mode	Zone-based automatic fan speed adjustment	V100R018C00SPC500, V100R019C1xSPC200, and respective later versions
	Automatic shutdown of unused high-speed interfaces	V100R018C00SPC500, V100R019C1xSPC200, and respective later versions
Deep mode	Zone-based automatic fan speed adjustment	V100R018C00SPC500, V100R019C1xSPC200, and respective later versions
	Automatic shutdown of unused high-speed interfaces	V100R018C00SPC500, V100R019C1xSPC200, and respective later versions
	Automatic shutdown of unused ports	V100R018C00SPC500, V100R019C1xSPC200, and respective later versions
	Automatic shutdown of unused boards	V100R018C00SPC500, V100R019C1xSPC200, and respective later versions
Optimal mode	Zone-based automatic fan speed adjustment	V100R018C00SPC500, V100R019C1xSPC200, and respective later versions
	Automatic shutdown of unused high-speed interfaces	V100R018C00SPC500, V100R019C1xSPC200, and respective later versions

 Copyright © Huawei Technologies Co., Ltd.

Energy Saving Mode	Energy Saving Measure	Supported Versions (Replace *x* in the Version Number with 0, 1, or 2.)
	Automatic shutdown of unused ports	V100R018C00SPC500, V100R019C1xSPC200, and respective later versions
	Automatic shutdown of unused boards	V100R018C00SPC500, V100R019C1xSPC200, and respective later versions
	Dynamic frequency adjustment	V100R018C00SPC500, V100R019C1xSPC200, and respective later versions
Custom mode	Zone-based automatic fan speed adjustment	V100R022C1xSPH203 and later versions
	Automatic shutdown of unused high-speed interfaces	V100R022C1xSPH203 and later versions
	Automatic shutdown of unused ports	V100R022C1xSPH203 and later versions
	Automatic shutdown of unused boards	V100R022C1xSPH203 and later versions
	Automatic shutdown of unused channels	V100R022C1xSPH203 and later versions
	Intelligent board scheduling	V100R022C1xSPH203 and later versions

 Copyright © Huawei Technologies Co., Ltd.

19.9.6.1.3 Energy Saving Parameter Conditions

Energy Saving Measure	Energy Saving Parameter	Supported Versions (Replace *x* in the Version Number with 0, 1, or 2.)
Automatic shutdown of unused boards	Manual shutdown	V100R022C1xSPC200 and later versions
	ONT auto-discovery check disabling	V100R022C1xSPC200 and later versions
Automatic shutdown of unused ports	Manual shutdown	V100R022C1xSPC200 and later versions

19.9.6.2 Board Conditions

19.9.6.2.1 Energy Saving Configuration Conditions

Due to board and subrack model differences, each energy saving measure only takes effect for certain boards. The following figure shows the conditions:

Energy Saving Measure	Applicable Hardware
Zone-based fan speed adjustment	All multi-zone subracks such as X17 and X7 (X15, X2, and X1 are single-zone subrack.)
Shutdown of unused high-speed interfaces	All service boards
Shutdown of unused boards	Supported by EPON/GPON/ETH boards Not supported by control, power, interface, or AEC boards Not supported by upstream or L-type boards Not supported by X1 subrack boards
Shutdown of unused ports	Supported by ports on EPON/GPON boards; not supported by upstream ports Only GPON, EPON, and Flex-PON boards in V100R022C10SPH211 version support power-off of optical modules; other boards do not support this function
Shutdown of unused channels	Supported by dual-channel boards such as Combo boards and 10E PON boards; not supported by L-type boards

 Copyright © Huawei Technologies Co., Ltd.

Energy Saving Measure	Applicable Hardware
Dynamic frequency adjustment	Boards using 5121 chips
Intelligent board scheduling	Boards using 5122 chips

You can query detailed board conditions through the following specification query tools:

Operators: **(Operator) Huawei Optical Product Info-Finder (huawei.com)**

Enterprise customers: **(Enterprise) Huawei Optical Product Info-Finder (huawei.com)**

19.9.6.2.2 Support for Energy Consumption Visualization

After enabling the energy saving function, you can enable the energy saving visualization function on the network management system (NMS). The energy saving visualization function depends on the capability of the power board. The following table lists the power boards that support energy saving visualization.

Subrack	Power Board Name	Whether Supporting Energy Saving Visualization	Start Version
X17/X15/X7	H901PILA	N	-
X17/X15/X7	H901PILB	N	-
X17/X15/X7	H901PIMA	N	-
X17/X15/X7	H902PILA	N	-
X17/X15/X7	H903PILA	Y	R22C01 R22C02
X17/X15/X7	H905PILA	Y	R22C01 R22C02
X2	H901PISA	N	-
X2	H901PISB	Y	R22C01 R22C02
X2	H901PISC	Y	R22C01 R22C02

 Copyright © Huawei Technologies Co., Ltd.

Subrack	Power Board Name	Whether Supporting Energy Saving Visualization	Start Version
X2	H902PISB	Y	R22C01 R22C02
X2	H903PISB	Y	R22C01 R22C02
X1	H931PICA	Y	R24C01 R24C02
X1	H931PICB	Y	R24C00 R24C01 R24C02
X1	PAC300S12	Y	R24C01 R24C02
X1	PDC300S12	Y	R24C01 R24C02

CAUTION

1. All power boards on a device must support the energy saving visualization function for the NMS to support energy saving visualization.

2. If a power board on a device is faulty, the energy saving visualization function may be abnormal.

19.9.7 FAQ

19.9.7.1 Can the Custom Energy Saving Mode Be Configured on a Device?

You need to apply for a license before configuring the custom energy saving mode. Currently, you can only switch to the custom mode on the NMS. The CLI does not support this operation.

19.9.7.2 Does a Custom Energy Saving Measure Take Effect Immediately After Configuration?

A custom energy saving measure only takes effect in custom mode. There are two scenarios:

1. If the system runs in non-custom mode (for example, the basic mode), a configured custom energy saving measure does not take effect. You need to switch to the custom mode for the measures to take effect.

 Copyright © Huawei Technologies Co., Ltd.

2. If the system runs in custom mode, a configured custom energy saving measure immediately takes effect after configuration.

19.9.7.3 Link for the Manual of Configuring Custom Energy Saving Measures on the NMS

Path: **iMaster NCE-FAN Product Documentation** > **NCE-FAN Application Operation and Maintenance** > **Access Controller** > **Energy Consumption Management**

V100R022C10 version link: **https://support.huawei.com/hedex/hdx.do?docid=DOC1101081994&id=EN-US_TOPIC_0000001321356616**

19.10 Version and Data Management

This topic provides an introduction to the version and data management feature, and describes the working principle of this sub feature.

NOTICE

After saving the data, keep the data properly to avoid security risks.

19.10.1 Introduction

Definition

Version and data management includes patch management, rollback function, configuration data management, and version upgrade.

Purpose

This sub feature facilitates carriers in version upgrade and maintenance.

Benefits

Benefits to carriers: The carriers' operating expenditure (OPEX) is saved considerably, and the customer satisfaction is increased.

19.10.2 Principle

Patch Management

The flash memory (storage medium in the system) has a patch area to store the loaded patches.

A patch can be a hot patch or cold patch. The system needs to be restarted for a cold patch to take effect or stop functioning. Nevertheless, in the case of a hot patch, the system need not be restarted for the same purpose. A hot patch supports the rollback function; therefore, the hot patch can be rolled back to the status before the latest hot patch is loaded.

 Copyright © Huawei Technologies Co., Ltd.

In addition, a patch can be activated, deactivated, run, or deleted. The loaded patch is deactivated by default; therefore, to make the loaded patch take effect, activate it. To make the patch take effect after the system restart, activate and run the patch before the system is restarted.

The system supports the following four types of patches:

- HP refers to the host hot patch. It takes effect after being loaded and then activated. For a user, this type of patches, after being loaded, is displayed as HPXXX.
- SPH is the set of HP patches. It takes effect after being loaded and activated. For a user, this type of patches, after being loaded, is displayed as SPHXXX, without displaying the status of HP patches
- CP refers to the host cold patch. It takes effect after it is loaded and the system is restarted. For a user, this type of patches, after being loaded, is displayed as CPXXX.
- SPC is the set of CP patches. It takes effect after being loaded and activated. For a user, this type of patches, after being loaded, is displayed as SPCXXX, without displaying the status of CP patches.

Rollback Function

The flash memory of the control board is divided into two same storage areas (namely, active storage area and standby storage area) to store the program, database, and extended BIOS. The storage area that is operating currently is the active storage area. When the program, database, and extended BIOS are upgraded, the new program, database, and extended BIOS are loaded to the standby storage area. After the system is restarted, the system automatically loads the new program, database, and extended BIOS. The rollback function is implemented based on two sets of program, database, and extended BIOS in both the active and standby storage areas.

By default, after upgrade, the system saves the pre-upgrade host program and database for 48 hours. 48 hours later, the system automatically cancels the rollback function. That is, 48 hours later, the system duplicates the program, database, and extended BIOS in the operating area to the standby storage area. In this way, the versions in both the active and standby storage areas are the same. You can run the **set norollback time** command to set the time for canceling the rollback function.

The system supports automatic rollback and manual rollback. After version upgrade, if the system fails to start up, the system is automatically rolled back to the version before upgrade. After version upgrade, if the system becomes abnormal during the running and cannot recover, you can run the **rollback** command to roll back the system to the version before upgrade.

Configuration Data Management

- Saving the configuration data manually: The current configuration data can be saved manually through the **save** commands. If the configuration data is not saved before the system is reset or restarted, it will be lost after the system reset or restart. Therefore, manually save the configuration data once before the system is reset or restarted.

 Copyright © Huawei Technologies Co., Ltd.

- Saving the configuration data after any changes to the configuration data: After the configuration data is changed, the system will save the changed configuration data automatically at a preset interval. You can run the **autosave interval** command to set the interval.
- Saving the configuration data at a preset time or interval: In the system, the configuration data can be saved automatically at a preset time or interval. This time or interval is user-defined. For example, the time or interval can be set to 23:00 or two hours respectively. In this case, the configuration data is saved at an interval of two hours or at 23:00.

The data erasure operation can be performed to restore the configuration data of the device to the default settings.

The system also supports backing up the current configuration data manually or at a preset time to a specified file server.

Version Upgrade

Software version in the system can be upgraded through the CLI or the NMS by using FTP/TFTP/XMODEM/SFTP.

NOTE

FTP, TFTP and XMODEM have security risks due to protocol restrictions. SFTP is recommended.

19.10.3 ISSU

Definition

In-service software upgrade (ISSU) is an upgrade mode in which services are not interrupted or interrupted for only a short period of time.

Benefits

With the development of fiber to the mobile base station (FTTM), fiber to the office (FTTO), and large-capacity fiber to the home (FTTH) services, customers require shorter or zero service interruption during an upgrade. The previous generation of OLT supports ISSU, but has the following problems:

- When control boards work in active/standby mode and service boards are not configured with inter-board type X protection switching, services are not interrupted during control board upgrade, but are interrupted when the service boards are reset.
- When control boards work in active/standby mode and service boards and upstream boards are configured with inter-board type X protection switching, services are interrupted for less than 200 ms during an upgrade. In this scenario, all these boards are in dual configurations, which increases operation costs.

The MA5800 enhances the ISSU feature. Specifically, when control boards work in active/standby mode and service boards are not configured with inter-board type X protection switching, ONUs remain online and services are interrupted for less than 10s during an upgrade. The enhanced ISSU feature improves system reliability and helps reduce operation costs.

 Copyright © Huawei Technologies Co., Ltd.

Principles

Prerequisite

- Control boards working in active/standby mode are configured.
- The upstream port is provided by the control board or a backup upstream board is configured, and inter-board protection or aggregation is configured.

Working Principles

The following figure shows the principles of MA5800 ISSU.

Figure 19-32 Working principles

ISSU principles

Boards of the MA5800 support decoupling of the control plane and forwarding plane, that is, the control plane and forwarding plane can be upgraded separately.

- When the control plane is being upgraded, the forwarding plane works continuously and therefore services generated before the upgrade are not affected.
- When forwarding entries of the forwarding plane are upgraded, services are interrupted for about 10s.

Through the separate upgrade of the control plane and forwarding plane, boards of the MA5800 ensure that ONUs are continuously connected, services generated

 Copyright © Huawei Technologies Co., Ltd.

before the upgrade are affected for only about 10s, and services generated during the upgrade are affected for about 2 minutes.

NOTE

Existing services refer to the services that have been provisioned before the ISSU upgrade. New services refer to the newly configured services that have not been provisioned before the ISSU upgrade.

ISSU Upgrade

An MA5800 ISSU upgrade consists of the following operations:

1. Perform a pre-upgrade check.
2. Save and back up the database file.
3. Load the combined package file.
4. Run the **upgrade issu** command to activate the package file.

   NOTE

   When MSTP is also configured on this board, ISSU does not take effect for this board. Services of this board are interrupted about 2-3 minutes and recover after the board resets.
5. Perform a post-upgrade check.

NOTE

For more details about the upgrade, see the Upgrade Guide released together with the product software. For its potential risks, it is recommended to perform an upgrade at night when the network carries a smaller volume of service traffic.

19.11 Anti-theft

Anti-theft for the MA5800 refers to using software to prevent the device from being misused by unauthorized users.

19.11.1 Anti-theft Feature Updates

Version	Description
V100R019C2X and later versions	First version to support this feature.

19.11.2 Anti-theft Introduction

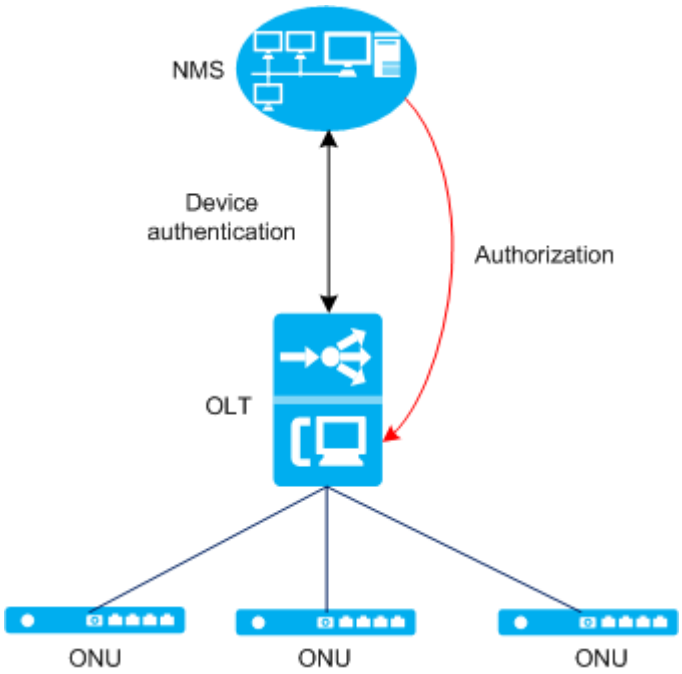
Anti-theft for the MA5800 is a technology that allows users to lock and unlock the device using public and private keys. The public key is loaded onto the device for enabling anti-theft and the private key is used to generate an authentication file for unlocking the device. In this way, the device is locked and can be used only if it is unlocked using a matched private key. Anti-theft effectively prevents unauthorized users from using devices.

- Control boards and service boards are authorized separately.

 Copyright © Huawei Technologies Co., Ltd.

- After a device is loaded with a public key (the anti-theft function is enabled) and powered on or restarted, it requests the NMS to perform private key authorization. After the authorization is successful, the device can be normally used.
- If a device to which the public key is loaded cannot obtain a private key from the NMS or the private key does not match, the device restriction policies are enabled.
- If a device is enabled with the anti-theft function but fails to be authorized, the device cannot be downgraded to a version that does not support the anti-theft function.

Figure 19-33 Anti-theft



19.11.3 Anti-theft Principles

After anti-theft is enabled on a hardware device, the functions of the device will be restricted upon unauthorized use of the device.

Anti-theft Modes

Anti-theft is based on an asymmetric encryption and decryption algorithm and the RSA public/private key algorithm. Users use the NMS or third-party software to generate two keys (public key and private key). The public key is like a lock and is loaded to the device that requires the anti-theft function. The private key is the key to the lock and is loaded to the NMS.

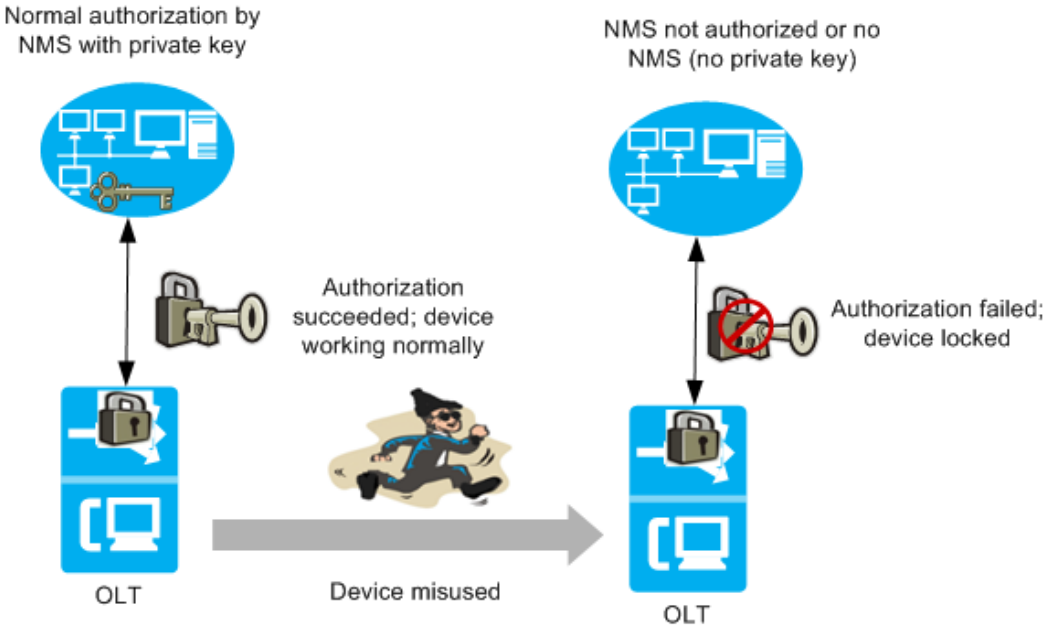
After a device is connected to the NMS and enabled with anti-theft, the NMS sends the public key to a device. After the main board is loaded, the anti-theft function is enabled. Any board that supports anti-theft automatically enables this function without additional configurations.

After a device to which the public key is loaded is powered on or restarted, the device instructs the NMS to perform private key authentication. After the

 Copyright © Huawei Technologies Co., Ltd.

authentication succeeds, the device can be normally used. If a device to which the public key is loaded cannot obtain the private key from the NMS or the private key does not match, the device is locked to prevent misuse, as shown in **Figure 19-34**.

Figure 19-34 Anti-theft principles



Anti-theft Mechanism

The anti-theft function restricts device usage using configuration and service control policies. If authorization fails after a device goes online or restarts, the configuration control policy takes effect immediately and the grace period countdown starts. The grace period can be customized and is set to 7 days by default. During the grace period, the service control policy does not take effect. After the grace period expires, the configuration and service control policies take effect at the same time.

Configuration Control Policy

If a device enabled with the anti-theft function cannot be authorized, the device is locked. If the device fails to obtain authorization, it enters the unauthorized state. After the grace period expires, the device enters the locked state.

- Device configurations can be viewed but cannot be modified. In addition, non-root users cannot log in to devices using commands.
- The upstream port works normally.

Service Control Policy

- If anti-theft is enabled on the control board, service boards follow the protection policy of the control board.
- If anti-theft is disabled on the control board or the authorization information of a service board is inconsistent with that of the control board, the service board enters the independent anti-theft control mode. After the grace period expires, the port bandwidth is limited.

 Copyright © Huawei Technologies Co., Ltd.

19.11.4 Configuring Anti-theft

A device needs to work with the NMS to implement anti-theft. For details about how to configure anti-theft, see OLT management in the NMS operation guide.

NOTICE

- If a device is restarted due to a power failure or other reasons and cannot be connected to the NMS for authentication, the device enters the locked state and needs to be handled in time; otherwise, services will be affected due to anti-theft after the grace period expires.
- To facilitate management, it is recommended that an operator use the same key or root certificate on the entire network. If different certificates are used, different boards cannot be used together.

19.12 Environment Monitoring

This document describes the environment monitoring function of the MA5800 series products.

Purpose

This document provides:

- Basic knowledge about environment monitoring, including monitoring principles, hardware configuration, monitoring solutions, and sensors
- Guidance on configuring deployment data
- Guidance on troubleshooting maintenance faults

By reading this document, you can have a comprehensive understanding of environment monitoring.

19.12.1 EMUB Monitoring Solution

In the EMUB monitoring solution, the environment monitoring unit type B (EMUB) converts monitoring parameters and reports them to the control system.

19.12.1.1 DCDU-12E PDB

A direct current distribution unit-12E (DCDU-12E) provides –48 V DC power for all components in a cabinet. When working with the EMUB to provide environment monitoring for the MA5800, the DCDU-12E provides power input for the EMUB.

Function

The DCDU-12E provides ten –48 V DC outputs for communications equipment.

Appearance

Figure 19-35 shows the exterior of a DCDU-12E.

 Copyright © Huawei Technologies Co., Ltd.

Figure 19-35 DCDU-12E power distribution box (PDB)

Ports

Figure 19-36 shows the ports on a DCDU-12E.

Figure 19-36 Ports on a DCDU-12E panel

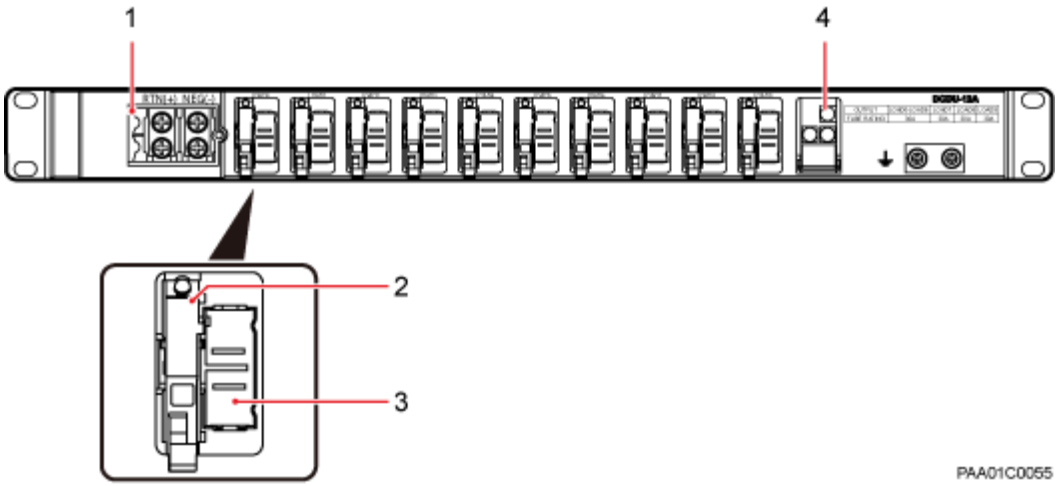


Table 19-17 lists the ports on a DCDU-12E panel.

Table 19-17 Ports on a DCDU-12E panel

No.	Port	Label	Connector and Cable	Description
(1)	DC input terminals	NEG(-)	Two one-hole OT terminals (M6). The default cross-sectional area of the cable is 16 mm^2.	Indicates the negative input terminal.
		RTN(+)		Indicates the positive input terminal.

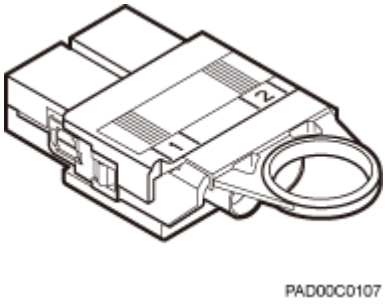
 Copyright © Huawei Technologies Co., Ltd.

No.	Port	Label	Connector and Cable	Description
(2)	Fuse block	LOAD0–LOAD9	-	The fuse block controls ports LOAD0–LOAD9, and thereby controls the power supplies to the BBU, fan assemblies, and transmission equipment. The indicator on the fuse block indicates the status of a fuse. • If the indicator is on, the fuse is faulty and needs to be replaced. • If the indicator is off, the fuse is working properly.
(3)	DC output terminals	LOAD0–LOAD9	Figure 19-37 shows an EPC4 connector which applies to ports LOAD0–LOAD9. The cross-sectional area of the cable supported ranges from 1.5–4 mm^2.	-
(4)	Spare fuse box	-	-	It contains three 30 A spare fuses.

NOTE

A fuse block and DC output terminals combined are called a fuse terminal block.

Figure 19-37 Exterior of an EPC4 connector



Specifications

Table 19-18 lists the technical specifications of a DCDU-12E.

 Copyright © Huawei Technologies Co., Ltd.

Table 19-18 Technical specifications of a DCDU-12E

Parameter	Specifications
Dimensions (H x W x D)	A DCDU-12E is 1 U high and can be installed in a 19-inch cabinet or rack. Its dimensions are as follows: • 42 mm x 442 mm x 65 mm (without mounting ears) • 42 mm x 482.6 mm x 65 mm (with mounting ears)
Input voltage/ current	Input voltage: –48 V DC Input current: 160 A
Port surge protection	Surge protection for a –48 V DC power port: • Differential mode: 2 kV (1.2/50 μs) • Common mode: 4 kV (1.2/50 μs)
	Surge protection for a –48 V DC power port • Differential mode: 10 kA (8/20 μs) • Common mode: 20 kA (8/20 μs)

19.12.1.2 EMUB Monitoring Unit

This section describes the appearance, functions, ports, indicators, DIP switches, and technical specifications of the EMUB.

Appearance

The EMUB is a 19-inch wide and 1 U high case-shaped device. **Figure 19-38** shows the EMUB.

Figure 19-38 EMUB appearance

Function

The EMUB is an environment monitoring device. It monitors and detects various input signals. **Table 19-19** describes the functions of the EMUB.

 Copyright © Huawei Technologies Co., Ltd.

Table 19-19 EMUB function description

Function	Description
Power input	• Provides one power supply input port and supports –48 V DC power input (–36 V DC to –72 V DC). • Supports input voltage detection.
Environment detection	• Provides an independent port for the +12 V DC/+24 V DC current-type (4–20 mA) temperature/humidity (T/H) sensor. The temperature detection range is –20°C to +80°C. • Provides four ports for monitoring different analog signals. Voltage-type (0–5 V) and current-type (4–20 mA) sensors can be connected. • The input type is configured by the base transceiver station (BTS). You can set the DIP switch to match the BTS configuration type. • Provides monitoring ports for the +24 V DC smoke sensor, +24 V DC (or +12 V DC) water sensor, +24 V DC (or +12 V DC) infrared sensor, and dry contact-type door status sensor. These ports can meet most monitoring requirements of the system.
Boolean signal detection	Provides 32 ports for monitoring Boolean signals. The Boolean signals can be transmitted over dry contacts or open collectors (OCs).
Relay node control	Provides six ports of the relay node type for controlling the output of external Boolean signals.
Communication function	Provides two RS-485 communications ports.

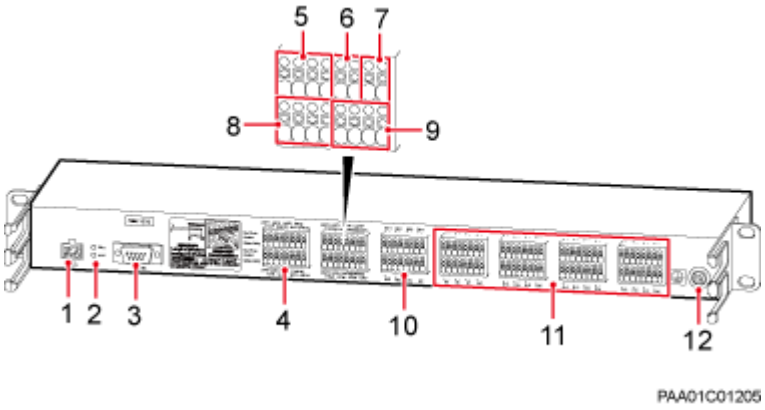
Ports Definition

The EMUB provides a power port, RS485 communications ports, detection ports, and output control ports.

Figure 19-39 shows the ports on an EMUB.

 Copyright © Huawei Technologies Co., Ltd.

Figure 19-39 Ports on an EMUB



(1) Power port	(2) Indicators	(3) RS485 port	(4) Analog signal port
(5) Infrared sensor port	(6) Smoke sensor port	(7) Door status sensor port	(8) T/H sensor port
(9) Water sensor port	(10) Relay node Boolean signal output control port	(11) Boolean signal port	(12) Ground terminal

Power port

Figure 19-40 shows the position of the power port on the EMUB.

Figure 19-40 Power port

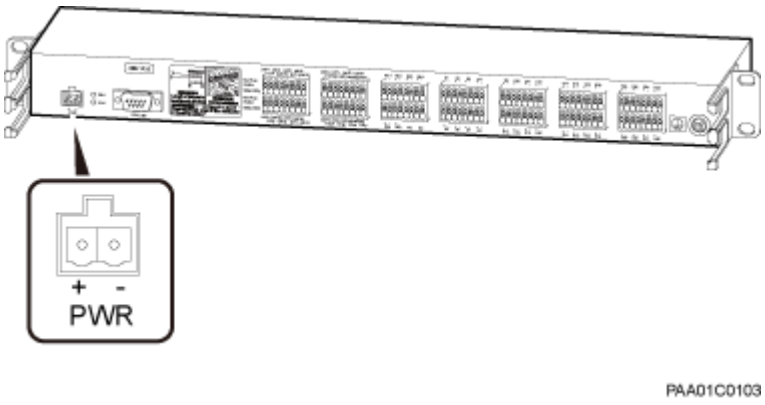


Table 19-20 describes the power port.

Table 19-20 Power port

Silk Screen	Connector Type	Description
PWR	2–pin connector	Power input port:

 Copyright © Huawei Technologies Co., Ltd.

Communications port

Figure 19-41 shows the position and socket pins of the RS485 port on the EMUB.

Figure 19-41 RS485 port

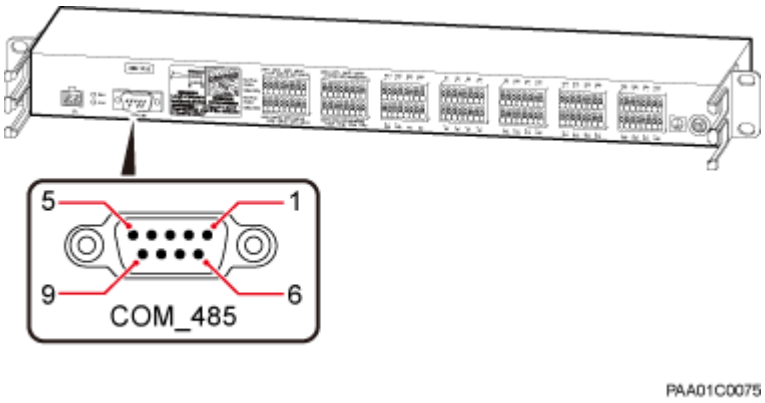


Table 19-21 RS485 port

Silk Screen	Connector Type	Description
COM_485	DB9 male connector	Communications port

Table 19-22 describes the pins of the RS485 port.

Table 19-22 RS485 port pins

PIN	Definition
1	NULL
2	RX1+: positive Rx end of port 1
3	TX1+: positive Tx end of port 1
4	RX2+: positive Rx end of port 2
5	TX2+: positive Tx end of port 2
6	RX1-: negative Rx end of port 1
7	TX1-: negative Tx end of port 1
8	RX2-: negative Rx end of port 2
9	TX2-: negative Tx end of port 2

 Copyright © Huawei Technologies Co., Ltd.

NOTE

- The Tx and Rx signals in the preceding table are defined in the direction from the EMUB.
- When the TXD+ is connected to the RXD+ and the TXD- is connected to the RXD-, the RS485 port can be used in two-wire mode.

Detection ports

The EMUB provides detection ports for the water sensor, smoke sensor, infrared sensor, door status sensor, and T/H sensor. It also provides input control detection ports for extended Boolean and analog signals.

Table 19-23 Detection ports on the EMUB

Signal Port	Silk Screen	Pin Function	Detected Signal Type
Water sensor port	24VW	+24 V DC power output	Voltage-type Boolean signal
	12VW	+12 V DC power output	
	WATER	Water sensor signal input	
	GND	Signal ground	
Smoke sensor port	24VS	+24 V DC power output	Current-type Boolean signal
	SMOKE	Smoke sensor signal input	
Infrared sensor port	24VO	24 V DC power output	Voltage-type Boolean signal
	12VO	12 V DC power output	
	OPTI	Infrared sensor signal input	

 Copyright © Huawei Technologies Co., Ltd.

Signal Port	Silk Screen	Pin Function	Detected Signal Type
	GND	Signal ground	
Door status sensor port	GATE+	Positive input end of door status signals	Dry contact Boolean signal
	GATE-	Negative input end of door status signals	
T/H sensor port	24VT	24 V DC power output	Current-type analog parameter
	TEMP	Temperature signal input	
	24VH	24 V DC power output	
	HUMI	Humidity signal input	
Boolean signal port	Si+ (i = 1–32)	Positive input end of Boolean signals from route i	Boolean signals in dry contact and OC modes Ports S13–S16 and ports S29–S32 support input Boolean signals of the voltage type. **Table 19-24** lists the valid Boolean signals of the voltage type.
	Si- (i = 1–32)	Negative input end of Boolean signals from route i	
Analog signal port	24Vi (i = 1–4)	24 V DC power output from route i	Current-type or voltage-type

 Copyright © Huawei Technologies Co., Ltd.

Signal Port	Silk Screen	Pin Function	Detected Signal Type
	12Vi (i = 1–4)	12 V DC power output from route i	
	ANAi (i = 1–4)	Analog signal input from route i	
	GNDi (i = 1–4)	Signal ground of route i	
Relay node Boolean signal output control port	REi+(i=1-6)	Positive output from Relay node of route i	Boolean signal
	REi-(i=1-6)	Negative output from Relay node of route i	
	NC	Reserved	

Table 19-24 Definition of valid voltage-type Boolean signals

Amplitude of input voltage signals	Detection Result
x ≤ 3 V DC	Low level
x ≥ 5 V DC	High level
3 V DC < x < 5 V DC	Uncertain

Indicators

Table 19-25 describes the indicators on an EMUB.

 Copyright © Huawei Technologies Co., Ltd.

Table 19-25 Indicators on the EMUB

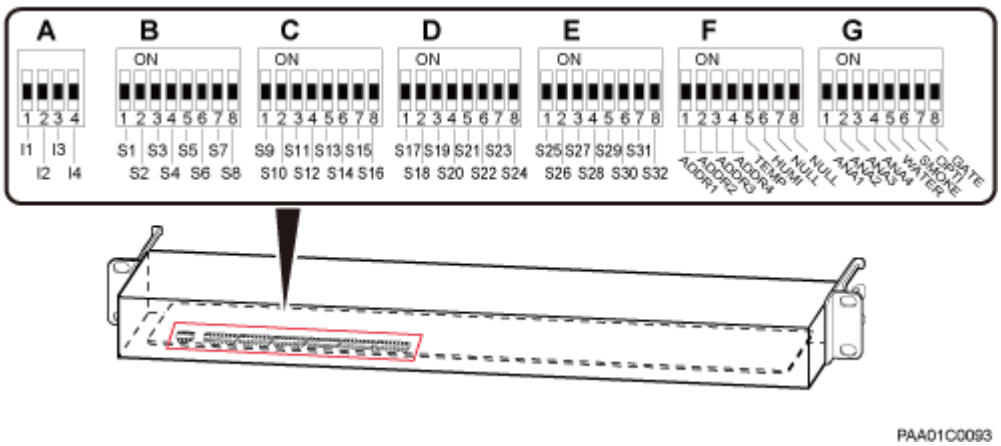
Label	Color	Status	Indication
RUN	Green	Steady off	No power is available.
		Blinking (on for 1s and off for 1s)	The board is registered and the communication is normal.
		Blinking (on for 0.125s and off for 0.125s)	The board is not registered or the communication is not set up.
ALM	Red	Steady off	The device is working properly.
		Steady on	An alarm is generated and the board needs to be replaced.
		Blinking (on for 1s and off for 1s)	An alarm is generated, and you need to locate the fault before determining whether to replace the board.

DIP Switches

This section describes the positions and functions of the DIP switches on the EMUB.

Figure 19-42 shows the positions of the DIP switches on the EMUB.

Figure 19-42 Positions of the DIP switches on the EMUB



 Copyright © Huawei Technologies Co., Ltd.

The ADDR DIP switch is used to select the communications protocol. It is located in position F in the preceding figure. The ADDR1-4 DIP switch is set to 0010 by default. 0 indicates that a toggle is set to ON, and 1 indicates that a toggle is set to OFF.

The DIP switches of sensors, Boolean signals, and analog signals are located in positions A, B, C, D, E, F, and G. **Table 19-26** describes functions of the DIP switches.

Table 19-26 DIP switches for sensors, Boolean signals, and analog signals

No.	DIP Switch	Function	Toggle State	Description
F(5)	TEMP	Bit 5: temperature sensor setting	ON	Used
			OFF	Not used
F(6)	HUMI	Bit 6: humidity sensor setting	ON	Used
			OFF	Not used
G(5)	WATER	Bit 5: water sensor setting	OFF regardless of whether it is used	
G(6)	SMOKE	Bit 6: smoke sensor setting	ON regardless of whether it is used	
G(7)	OPTI	Bit 7: infrared sensor setting	ON	Not used
			OFF	Used
G(8)	GATE	Bit 8: door status sensor setting	ON	Used
			OFF	Not used
G(1234)	ANA1 ANA2 ANA3 ANA4	Four analog sensor settings. Each bit corresponds to one analog sensor.	ON	Used
			OFF	Not used
A(1234)	I1 I2 I3 I4	Four analog sensor settings. Each bit corresponds to one analog sensor.	ON	Current type
			OFF	Voltage type

Copyright © Huawei Technologies Co., Ltd.

No.	DIP Switch	Function	Toggle State	Description
B. C, D, E	S1-S32	32 Boolean signal inputs S1–S32 corresponds to input signals 1 to 32.	ON	High level alarm
			OFF	Low level alarm or not used

Specifications

Table 19-27 and Table 19-28 and list the technical specifications of the EMUB.

Table 19-27 EMUB technical specifications

Item	Measurement Object	Specification
Input voltage detection	Measurement range	36–72 V DC
	Measurement precision	±1 V
Temperature measurement	Measurement range	–20°C to 80°C
	Measurement precision	±2°C
Humidity measurement	Measurement range	0%–100% RH
	Measurement precision	±5%
Boolean signal detection	Measurement content	Smoke, door status, water, and infrared sensors
	Alarm response time	< 10s
Analog signal detection	Measurement range	4–20 mA or 0–5 V
	Measurement precision	±1%
Relay output control	Control response time	<1s
	Dry-contact output capacity	Voltage: 110 V DC, 125 V AC, maximum load current: 1 A (30 V DC)
Communications rate	RS485	9600 bit/s

 Copyright © Huawei Technologies Co., Ltd.

Table 19-28 Performance specifications of the EMUB

Parameter	Specifications
Storage temperature	–40°C to +70°C (non-condensing)
Operating temperature	–20°C to +55°C
Storage relative humidity	5%–95% RH
Working humidity	10–90% RH
Working voltage type	DC
Rated operating voltage	–48 V
Rated working current	0.1 A
Dimensions	442 mm x 110 mm x 41.6 mm
Weight	About 1.6 kg
RS485 protection	• 3 kA in differential mode • 5 kA in common mode
Power supply protection	• 3 kA in differential mode • 5 kA in common mode
Dry contact protection	• 3 kA in differential mode • 5 kA in common mode
Dust-proof	IP20

19.12.1.3 Checking the Environment Monitoring Cables of the EMUB

The EMUB monitors the environment monitoring parameters collected by sensors in real time, and reports the alarm information to the control system. You need to check whether the connections from the EMUB monitoring unit to the control system are correct.

Figure 19-43 shows the connections of the EMUB environment monitoring cables, and **Table 19-29** describes the cable connections. By referring to the cable connection diagram, check whether the connections from sensors to the EMUB monitoring unit and from the EMUB monitoring unit to the control system are correct.

 Copyright © Huawei Technologies Co., Ltd.

Figure 19-43 Connection diagram for the environment monitoring cables of the EMUB

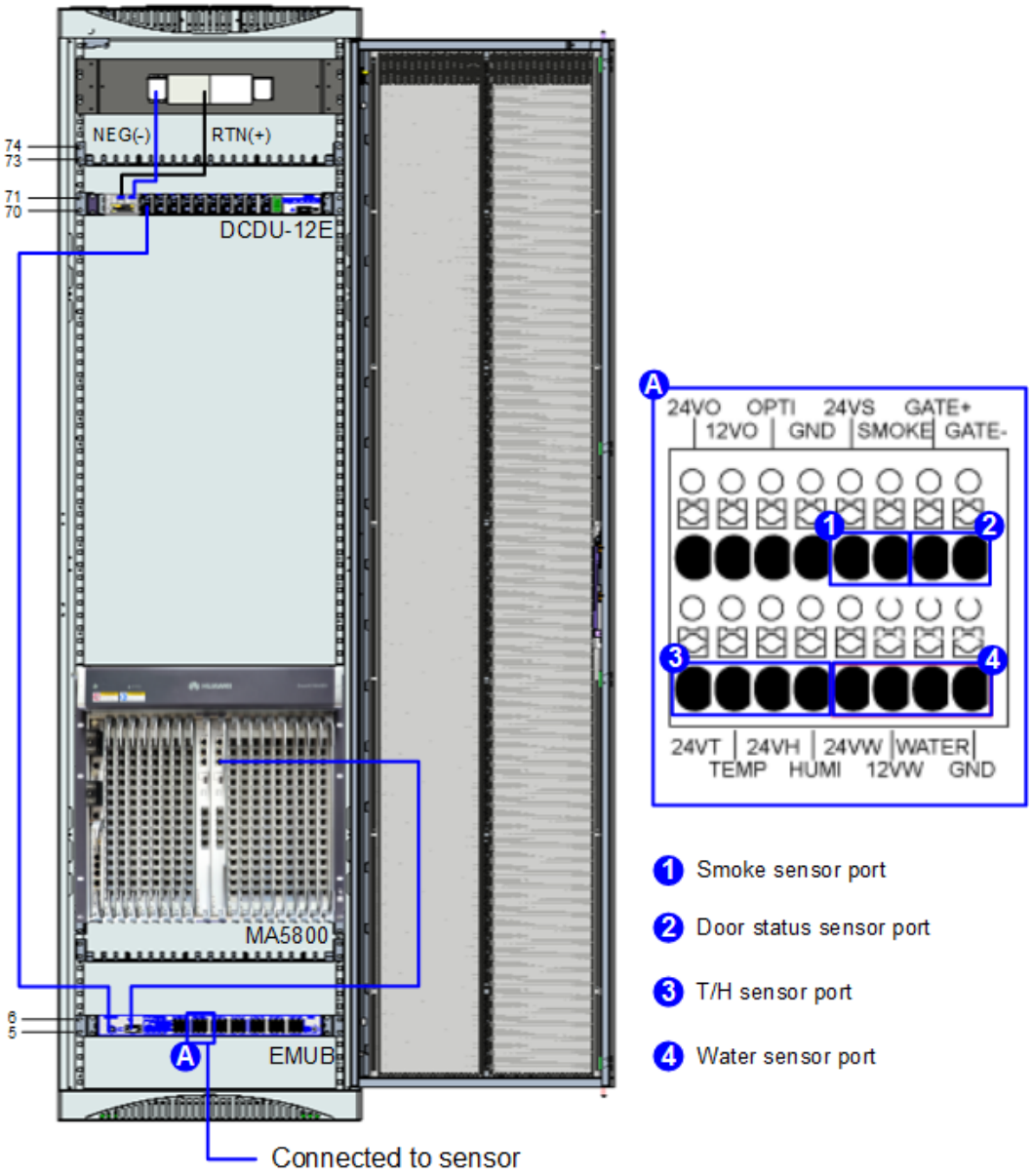


Table 19-29 Connection description for environment monitoring cables of the EMUB

Cable Name	From	To
Power cable from the cabinet PDB to the DCDU-12E	Output of the cabinet PDB	Input terminals on the DCDU-12E **NOTE** Only one power cable is required between the cabinet PDB and the DCDU PDB.

 Copyright © Huawei Technologies Co., Ltd.

Cable Name	From	To
Power cable from the EMUB to the DCDU-12E	Power port on the EMUB	Output terminals of the DCDU-12E
Monitoring cable from the EMUB to the control board	RS485 port on the EMUB	ESC port on the control board
Monitoring cable of the smoke sensor	24 VS/SMOKE of EMUB	Smoke sensor
Monitoring cable of the door status sensor	GATE+/GATE- ports on the EMUB	Door status sensor
Monitoring cable of T/H sensor	24VT/TEMP/24VH/HUMI of the EMUB	T/H sensor
Monitoring cable of the water sensor	24VW/12VW/WATER/GND of the EMUB	Water sensor

19.12.1.4 Commissioning the EMUB

This section describes how to commission the environment monitoring unit (EMUB) to ensure that it can monitor the environment status of the device.

Context

The EMUB uses various sensors to monitor environment parameters such as the ambient temperature, humidity, smoke, water, and door status.

When commissioning the EMUB, pay attention to the following points:

- The value range of the EMU subnode is 0–31.
- When the system is configured with multiple EMUs simultaneously, make sure that all sub-nodes are unique.

The EMUB supports the following analog and digital parameters:

- Analog parameters 0–3: Fixedly allocated by the system to –48 V power supply, 24 V power supply, and temperature and humidity respectively. The analog parameter names cannot be changed. Analog parameter 1 is reserved internally and cannot be used. For other analog parameters, only the upper and lower alarm thresholds and the monitoring switch can be configured.
- Analog parameters 4–7: User-defined analog parameters to be allocated to other extended analog sensors.
- Digital parameters 0–3: Fixedly allocated by the system to the water sensor, smoke sensor, infrared sensor, and door status sensor respectively. The names

 Copyright © Huawei Technologies Co., Ltd.

of these digital parameters cannot be changed. Digital parameter 2 is reserved internally and cannot be used.

- Digital parameters 4–35: user-defined digital parameters to be allocated to other extended digital sensors.

Procedure

Step 1 Run the **emu add** command to add an EMUB.

Step 2 Run the **interface emu** command to enter EMUB mode.

Step 3 Run the **emub analog** command to configure the upper and lower alarm and measurement thresholds for analog parameters to be monitored by the device in real time.

Step 4 Run the **emub digital** command to set the valid level, name, and alarm index of digital parameters of the EMU.

Step 5 Run the **save** command to save the data.

----End

Example

To add an EMUB, configure the analog and digital parameters of the EMUB, and then save the data, do as follows:

Table 19-30 shows the data planning.

Table 19-30 Data planning for configuring the EMUB

Item	Parameter	Remarks
EMU	Type: EMUB	The EMUB is displayed as EMUB in the command line.
	No.: 3	-
	Slave node ID: 2	-
Analog parameters	Analog parameter ID: 2	Fixedly allocated by the system to monitor the temperature. Only the upper and lower alarm thresholds and the monitoring switch can be configured. For the value and details of this parameter, refer to the **emub analog** command.
	Upper alarm threshold of analog parameter 2: 68°C	When the ambient temperature exceeds 68°C, the device reports an alarm.
	Lower alarm threshold of analog parameter 2: –5°C	When the ambient temperature is lower than –5°C, the device reports an alarm.

 Copyright © Huawei Technologies Co., Ltd.

Item	Parameter	Remarks
	Analog parameter ID: 3	Fixedly allocated by the system to monitor the humidity. Only the upper and lower alarm thresholds and the monitoring switch can be configured. For the value and details of this parameter, refer to the **emub analog** command.
	Upper alarm threshold of analog parameter 3: 80% RH	When the ambient humidity exceeds 80% RH, the device reports an alarm.
	Lower alarm threshold of analog parameter 3: 5% RH	When the ambient humidity is lower than 5% RH, the device reports an alarm.
Digital parameters	Digital parameter ID: 0	Fixedly allocated by the system to monitor the water sensor. Users can only modify the valid level of the digital parameter. Other parameters cannot be modified.
	Valid level of digital parameter 1: low level	The low level is used to represent the valid level. When the actual level of the monitored digital parameter is inconsistent with the valid level configured in the system, the device reports an alarm.
	Digital parameter ID: 1	Fixedly allocated by the system to monitor the smoke sensor. Users can only modify the valid level of the digital parameter. Other parameters cannot be modified.
	Valid level of digital parameter 1: high level	The high level is used to represent the valid level. When the actual level of the monitored digital parameter is inconsistent with the valid level configured in the system, the device reports an alarm.
	Digital parameter ID: 3	Fixedly allocated by the system to monitor the door status sensor. Users can only modify the valid level of the digital parameter. Other parameters cannot be modified.
	Valid level of digital parameter 2: high level	The high level is used to represent the valid level. When the cabinet door is opened, the device reports an alarm.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#emu add 3 EMUB 0 2 EMUB
huawei(config)#interface emu 3
huawei(config-if-emub-3)#emub analog 2 alarm-upper-limit 68 alarm-lower-limit -5
huawei(config-if-emub-3)#emub analog 3 alarm-upper-limit 80 alarm-lower-limit 5
huawei(config-if-emub-3)#emub digital 0 available-level low-level
huawei(config-if-emub-3)#emub digital 1 available-level high-level
huawei(config-if-emub-3)#emub digital 3 available-level high-level
huawei(config-if-emub-3)#display emub system parameter
huawei(config-if-emub-3)#quit
huawei(config)#save
```

19.12.2 Power3000 Monitoring Solution

The Power3000 monitoring solution is implemented by the CPMU01 monitoring module in the Power3000 power system. In this solution, the CPMU01 monitoring module reports the monitored parameters to the control system.

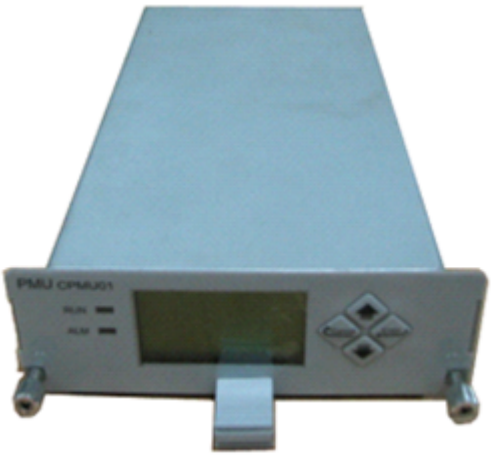
19.12.2.1 Monitoring Module CPMU01 of the Power3000

The CPMU01 of the Power3000 implements functions such as environment monitoring, abnormality alarm reporting, battery management, rectifier module management and power distribution management by using the local device, NMS server or hand-held terminal.

The CPMU01 displays real-time information (including operation parameters, operation status, alarm status, configuration parameters and control parameters) about the DC, AC, module and battery in the power system on the (liquid crystal display) LCD. Connected to the upper device by using the RS-232 and RS-485 serial ports on the monitoring backplane, the CPMU01 remotely manages the Power3000 system.

Figure 19-44 shows the appearance of the CPMU01.

Figure 19-44 Appearance of the CPMU01



 Copyright © Huawei Technologies Co., Ltd.

19.12.2.2 Checking the Environment Monitoring Cable of the Power3000

The CPMU01 is connected to the control system by using the RJ-45 serial port on the monitoring backplane. In the checking, ensure that the cable between the CPMU01 and the control system is connected properly.

Figure 19-45 shows the connection of the environment monitoring cable of the Power3000, and **Table 19-31** describes the cable connection. You can check whether the cable between the CPMU01 and the control system is connected properly based on the connection diagram.

Figure 19-45 Connection diagram of the environment monitoring cable of the Power3000

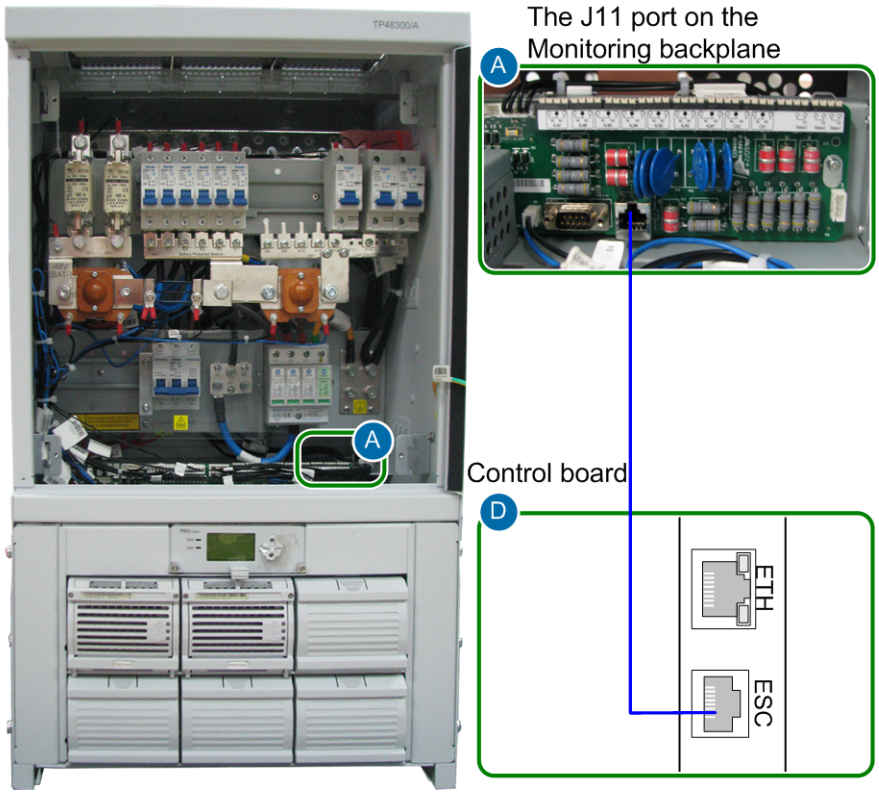


Table 19-31 Connection of the environment monitoring cable of the Power3000

Cable	Connection Position 1	Connection Position 2
Environment monitoring cable between the Power3000 and the control system	J11 port on the monitoring backplane of the Power3000	ESC port on the control board

 Copyright © Huawei Technologies Co., Ltd.

19.12.2.3 Configuring the Power3000 Environment Monitoring Mode

This topic describes how to configure environment monitoring parameters by using the CLI.

Configuration Process

The monitoring parameters can be reported to the control system only when the data for the Power3000 is configured correctly in the system.

Table 19-32 lists the commands used during the configuration.

Table 19-32 Commands for configuring the Power3000

To...	Run the Command...
Add an EMU	**emu add**
Query the configuration parameters of the power system	**display power system parameter**
Query the power environment parameters	**display power environment parameter**
Configure the battery charging parameters	**power charge**
Configure the battery management parameters	**power battery parameter**
Configure the temperature compensation coefficient of the battery	**power battery temperature**
Configure the battery set power-off parameters	**power off**
Configure the power distribution parameters	**power supply-parameter**
Configure the battery discharging test parameters	**power battery-test**
Configure the parameters for the power rectifier module	**power module-num**
Configure the battery high-temperature power-off parameters	**power temperature-off**
Configure the environment monitoring parameters	**power environment**
Configure the external extended digital parameters	**power outside-digital**

 Copyright © Huawei Technologies Co., Ltd.

1. Log in to the device through the maintenance terminal and add an EMU.
2. Query the status of the Power3000.
3. Enter the environment monitoring configuration mode and query the default configuration.
4. Configure the battery management parameters.
5. Configure the battery discharging test parameters.
6. Configure the battery high-temperature power-off parameters.
7. Configure the environment parameters.
8. Configure the extended digital parameters.
9. Query the information about the configured parameters and environment parameters of the power system.
10. Save the data.

Data Plan

Table 19-33 provides the data plan for configuring the monitoring parameters of the Power3000.

Table 19-33 Data plan for configuring the monitoring parameters of the Power3000

Item	Data Plan
EMU	Type: POWER3000 SN: 0 Subnode ID: 0 The subnode ID must be the same as the subnode setting of the corresponding DIP switches on the EMU, but the subnode ID must be different from IDs of the other subnodes on the same bus.
Charging parameters of the battery	Charging mode of the battery: automatic Equalized charging voltage of the battery: 56.5 V Float charging voltage of the battery: 53.5 V
Battery management parameters	Current-limiting coefficient for battery charging: 0.15 Interval of battery equalized charging: 60 days Number of battery sets: 1 Capacity of the battery set: 65 AH The battery capacity is configured according to the actual value.
Temperature compensation parameter of the battery	Upper temperature threshold of the battery set: 80°C Lower temperature threshold of the battery set: -20°C Temperature compensation coefficient of the battery set: 80 mV

 Copyright © Huawei Technologies Co., Ltd.

Item	Data Plan
Battery set power-off parameters	Battery set power-off permission status: permit Battery set power-off voltage: 43 V
Power distribution parameters	AC overvoltage alarm threshold of the power supply: 280 V AC undervoltage alarm threshold of the power supply: 180 V DC overvoltage alarm threshold of the power supply: 58 V DC undervoltage alarm threshold of the power supply: 45 V
Battery discharging test parameters	Battery auto-discharging test: auto-test Battery set auto-discharging test permission status: permit Battery auto-discharging test period: 60d
Power rectifier module parameters	The number of the power rectifier modules: 0 The number of the power rectifier modules is configured according to the actual value. POWER3000 supports up to 12 rectifier modules.
Battery high-temperature power-off parameters	Battery high-temperature power-off permission status: permit Temperature for battery high-temperature power-off: 53°C
Environment monitoring parameters	Upper alarm threshold of the temperature: 63°C Lower alarm threshold of the temperature: -5°C Upper test threshold of the humidity: 80°C Lower test threshold of the humidity: -20°C
External extended digital parameters	Digital parameter ID: 0-6 This digital parameter is set according to the actual requirements.

NOTE

The preceding data is configured according to the actual requirements. When the actually planned value of a parameter is the same as the default value, it is not necessary to configure the parameter.

Configuration Example

The following considers the configuration plan of the Power3000 as an example to describe the process of configuring the environment monitoring parameters of the Power3000 after connected.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#display emu 0
huawei(config)#interface emu 0
huawei(config-if-power3000-0)#display power system parameter
huawei(config-if-power3000-0)#display power environment parameter
huawei(config-if-power3000-0)#power battery parameter 1 0.15 60 65
huawei(config-if-power3000-0)#power battery-test auto-test permit 60
huawei(config-if-power3000-0)#power temperature-off battery-off-state permit battery-off-temperature 53
huawei(config-if-power3000-0)#power environment temperature 63 -5 80 -20
huawei(config-if-power3000-0)#power  outside-digital 0 digital-alarm 9 available-level  high-level name  door
huawei(config-if-power3000-0)#display power system parameter
huawei(config-if-power3000-0)#display power environment parameter
huawei(config-if-power3000-0)#quit
huawei(config)#save
```

19.12.3 ETP4890 Monitoring Solution (AC Power Supply)

The ETP4890 monitoring solution is implemented by the SMU01B monitoring module in the ETP4890 power system. In this solution, the SMU01B monitoring module reports the monitored parameters reported by the sensor transfer box to the control system.

19.12.3.1 ETP4890 Power System

An AC-powered cabinet uses the ETP4890 power system to convert the AC power to the DC power for power distribution. The small and modular ETP4890 power system supports easy installation and maintenance and effective LCD interface management. The power system supports a maximum of 90 A output current.

Function

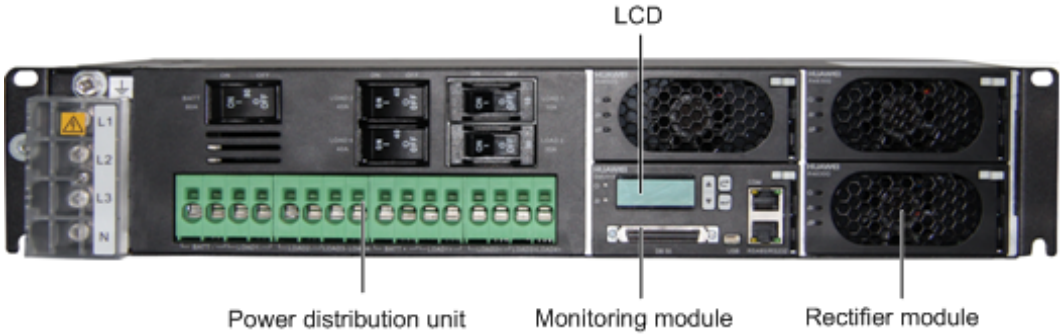
- Converts AC power into stable DC power for communications equipment.
- Communicates with the upper device using a COM or an RS485/RS232 serial port; sends control signals based on the values of monitored parameters or the control command sent by the upper device; detects power distribution parameters, power supply module parameters, and auxiliary digital parameters in real time.
- Connects to one to two groups of lead acid batteries and uses a monitoring module to manage the batteries; provides complete management for battery charging and discharging, ensuring efficient battery usage.
- Connects to a sensor transfer box using a DB50 port, detecting analog parameters and digital parameters, and outputting control signals.
- Uses a monitoring module to monitor all running parameters for the ETP4890 power system in real time; determines the working status; reports alarms in a timely manner.
- Clearly displays the monitoring and alarm information on the liquid crystal display (LCD) of the monitoring module of the ETP4890; allows operations such as parameter setting on the LCD of the monitoring module.

Appearance

Figure 19-46 shows the appearance of the ETP4890 power system.

 Copyright © Huawei Technologies Co., Ltd.

Figure 19-46 ETP4890 power system



Configuration

Table 19-34 lists the components of the ETP4890 power system.

Table 19-34 Component configuration of the ETP4890 power system

Component	Quantity
AC and DC power distribution subrack	1
Monitoring module	1
Power supply module R4830N1/ R4830G1	1 to 3 (optional; by default, the R4830N1 is configured)

The ETP4890 power system can be configured with three power supply modules, which are connected in parallel for output.

Table 19-35 shows the mapping between the quantity of power supply modules in the ETP4890 power system and the maximum output current.

Table 19-35 Mapping between the quantity of power supply modules of the ETP4890 power system and the maximum output current

Quantity of Power Supply Modules	Maximum Output Current
1	• 30 A (at 55°C) • 25 A (at 60°C)
2	• 60 A (at 55°C) • 50 A (at 60°C)
3	• 90 A (at 55°C) • 75 A (at 60°C)

 Copyright © Huawei Technologies Co., Ltd.

Communication Ports on the Monitoring Module

Table 19-36 lists the communication ports on the monitoring module. A user can set the baud rate of the device using the LCD interface of the monitoring module.

Table 19-36 Communication ports on the monitoring module

Communication Port	Communication Mode	Communication Parameter	Description
COM	RS485/RS232	Baud rate: 9600 bit/s or 19200 bit/s	Used as a serial port, communicating with the upper device
RS485/RS232	RS485/RS232	Baud rate: 9600 bit/s or 19200 bit/s	Used as a serial port, supporting the northbound/ southbound communications

NOTE

By default, the ETP4890 power system is equipped with a monitoring module, which is capable of the inband management (SMU01B). The COM port cannot be used as a network port.

Indicators

Table 19-37 lists the indicators on the power supply module of the ETP4890 power system.

Table 19-37 Indicators on the power supply module of the ETP4890 power system

Indicator	Color	Status	Description
Power indicator	Green	Steady on	Normal
		Blinking at a frequency of 0.5 Hz	There is AC input, and the power supply module is in the manual query state.
		Blinking at a frequency of 4 Hz	Data is being loaded online.
		Off	There is no AC input.

 Copyright © Huawei Technologies Co., Ltd.

Indicator	Color	Status	Description
Alarm indicator	Yellow	Steady on	A protection mechanism is triggered. After the fault is rectified, the power supply module can be restored to the normal state.
		Blinking	Communication between the power supply module and the site monitoring module is interrupted.
		Off	Normal
Fault indicator	Red	Steady on	An unrecoverable fault occurs.
		Blinking	The power supply module equipped with the EMU does not match the actual power supply module.
		Off	Normal

Table 19-38 lists the indicators on the monitoring module.

Table 19-38 Indicators on the monitoring module

Indicator	Color	Status	Description
RUN	Green	Off	The monitoring module is faulty or there is no DC power input.
		Blinking at a frequency of 0.5 Hz	The monitoring module is working properly and communicating with the upper device properly.

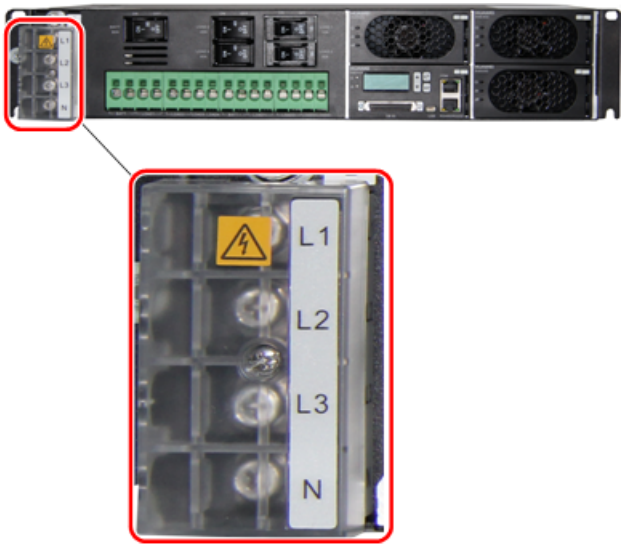
 Copyright © Huawei Technologies Co., Ltd.

Indicator	Color	Status	Description
		Blinking at a frequency of 4 Hz	The monitoring module works properly but does not communicate with the upper device properly.
ALM	Red	Off	The system does not generate any critical alarms.
		Steady on	The system generates critical alarms.

Input Terminal

The ETP4890 power system supports 220 V AC input. **Figure 19-47** shows the input terminals in the ETP4890 power system.

Figure 19-47 Input terminals in the ETP4890 power system



Output Terminal

The ETP4890 power system supports eight groups of output terminals. The left-side eight output terminals are negative and the right-side eight output terminals are positive.

Figure 19-48 shows the output terminals in the ETP4890 power system.

 Copyright © Huawei Technologies Co., Ltd.

Figure 19-48 Output terminals in the ETP4890 power system

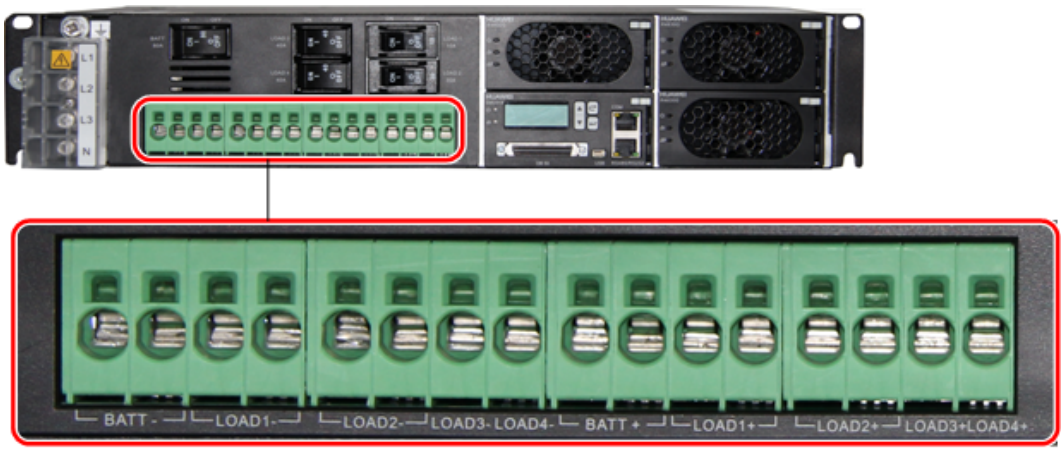


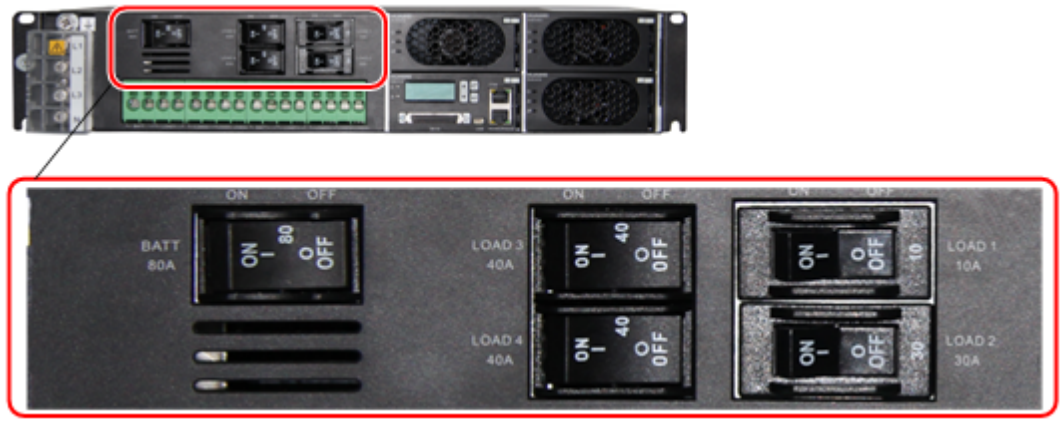
Table 19-39 lists the Specifications of the output terminals.

Table 19-39 Specifications of the output terminals

Port	Circuit Breaker
LOAD1	10 A
LOAD2	30 A
LOAD3	40 A
LOAD4	40 A
BATT	80 A

Figure 19-49 shows the output branch circuit breaker in the ETP4890 system.

Figure 19-49 Output branch circuit breaker in the ETP4890 system



Specifications

Table 19-40 lists the specifications of the ETP4890 power system.

 Copyright © Huawei Technologies Co., Ltd.

Table 19-40 Specifications of the ETP4890 power system

Specification Class	Specification	Description
Environment conditions	Operating temperature	-40°C to +70°C **NOTE** The power system can work at a temperature of 70°C for 8 hours.
	Operating humidity	5% to 95% (no condensation)
	Altitude	0 m to 4000 m (Note that in the range of 2000 m to 4000 m, the operating temperature decreases by 1°C when the altitude increases by 200 m.)
AC input	Mode	mono-phase, dual-live wire
	Voltage	85 V AC to 290 V AC (rated value: 110 V/220 V)
	Frequency	45 Hz to 66 Hz (rated value: 50 Hz)
	Power factor	≥ 0.99 in the case of a rated load
	Power supply module peak efficiency	• R4830N1: 94% • R4830G1: 96%
DC output	Voltage	-42 V to -58 V
	Default voltage	-53.5 V
	Maximum power	• R4830N1: 4815 W (input voltage: 176 V AC to 290 V AC; if the temperature is higher than 55°C, power decreases linearly); 2407.5 W (input voltage: 85 V AC to 175 V AC; if the temperature is higher than 55°C, power decreases linearly) • R4830G1: 4800 W (input voltage: 176 V AC to 290 V AC; if the temperature is higher than 55°C, power decreases linearly); 2400 W (input voltage: 85 V AC to 175 V AC; if the temperature is higher than 55°C, power decreases linearly) **NOTE** Maximum output power of the system = Number of the configured power supply modules x Output power of each power supply module

 Copyright © Huawei Technologies Co., Ltd.

Specification Class	Specification	Description
	Regulated voltage precision	≤ ±0.6%
	Peak noise voltage	≤ 200 mV (0 MHz to 20 MHz)
Structure	Dimensions (H x W x D) without brackets	86.1 mm x 442 mm x 255 mm
	Weight	≤ 10 kg (including three 30 A power supply modules and one SMU01B monitoring module)
	Protection class	IP20
	Installation mode	Installed in an IEC cabinet or an ETSI cabinet with brackets
	Maintenance mode	Front access
	Cooling method	Air cooling, with the built-in fan in the power supply module

19.12.3.2 PMIB02 Sensor Transfer Box

This topic describes the application and front panel of the sensor transfer box, and provides the specifications of the sensor transfer box.

Function

The sensor transfer box collects the signals of each sensor and forwards the signals to the monitoring module.

Overview

The sensor transfer box is an optional part. It provides diverse Sensor ports to connect the monitoring unit to the external sensors.

Front Panel

Figure 19-50 shows the PMIB02 sensor transfer box.

Figure 19-50 Front panel of the PMIB02 sensor transfer box

 Copyright © Huawei Technologies Co., Ltd.

Sensor Port

Figure 19-51 shows the ports on the PMIB02 sensor transfer box, and Table 19-41 describes the functions.

Figure 19-51 Sensor ports of the PMIB02 sensor transfer box

Table 19-41 Pin functions of the sensor ports on the sensor transfer box PMIB02

Silk Screen	Signal Port	Pin Function	Remarks
BAT_WE	Sensor port of battery temperature 1	• Pin 1 is for grounding. • Pin 2 is for the signal cable. • Pin 3 is reserved. • Pin 4 is for the +12 V power supply. The external sensor outputs the 0-5 V voltage signal.	Used
VBTEM2	Sensor port of battery temperature 2	• Pin 1 is for the +24 V power supply. • Pin 2 is for the temperature signal cable. The external sensor outputs the 4-20 mA current signal.	-
WATER	Water sensor port	• Pin 1 is for the +12 V power supply. • Pin 2 is for the signal cable. • Pin 3 is for grounding. • Pin 4 is reserved. When the impedance between pin 2 and pin 3 drops to a certain range, alarms are triggered.	-
TEM_HU	Sensor port of temperature and humidity 1	• Pin 1 and pin 3 are for the +24 V power supply. • Pin 2 is for the temperature signal cable. • Pin 4 is for the humidity signal cable.	Used

 Copyright © Huawei Technologies Co., Ltd.

Silk Screen	Signal Port	Pin Function	Remarks
VTEM2	Sensor port of temperature and humidity 2	• Pin 1 is for the +24 V power supply. • Pin 2 is for the temperature or humidity signal cable. The external sensor outputs the 4-20 mA current signal.	-
JTP1	Sensor port of the MDF	• Pin 1 is for the signal cable. • Pin 2 is reserved.	Used
SMOKE	Smoke sensor port	• Pin 1 is for the signal cable. • Pin 2 is for the +24 V power supply.	-
JTM1	Door-status sensor port	• Pin 1 is for the signal cable. • Pin 2 is for the +24 V power supply. By default, when pin 1 and pin 2 are connected, the door is in the normal state; when they are disconnected, alarms are triggered. If no door-status sensor is used, connect pin 1 and pin 2 to disable door-status alarms.	Used
JK1, JK2	Alarm digital ports	The maximum voltage is 60 V DC and the maximum load current is 500 mA. • When minor alarms are generated, JK2 is short-circuited. • When critical alarms are generated, JK1 is short-circuited. • Users can define alarm levels through the BAM program.	Reserved
JKM1-JKM4	Digital ports	The maximum voltage is 5.6 V DC and the maximum current is 20 mA.	-
SIM1, SIM2	Battery unbalanced sensor ports	• Pin 1 is for the signal cable. • Pin 2 is reserved.	-
JAC1-JAC6	Output ports of the optical coupler	Pin 1 and pin 2 are connected to pole C and pole E of the optical coupler. The optical coupling parameter Vce max is 40 V DC and Ic max is 80 mA. Avoid setting the parameters to the maximum values.	Reserved
FU_ALM	Detecting port of the battery fuse	• Pin 1 is for the signal cable FU1+ and is connected to the negative pole of the battery. • Pin 2 is reserved for the FU1- signal.	-

 Copyright © Huawei Technologies Co., Ltd.

Silk Screen	Signal Port	Pin Function	Remarks
JTD1-JTD7	Input ports of the standby Boolean value	• Pin 1 is for the +24 V power supply. • Pin 2 is for the +12 V power supply. • Pin 3 is for the signal cable. • Pin 4 is for grounding. According to the actual requirements, access related passive digital signals to pins 2 and pin 3.	Reserved
J1	DB50 port	A monitoring board and monitoring transfer board are connected through DB50 to input sensor signals and the output control signals.	-

NOTE

"JTD1-JTD7" are the input ports of the standby Boolean value. By default, the sensors are not configured. You can determine the type of signal to be accessed. With related configuration, the monitor unit can monitor them.

Specifications

Table 19-42 lists the specifications of the PMIB02 sensor transfer box.

Table 19-42 Specifications of the PMIB02 sensor transfer box

Sensor Transfer Box	Dimensions (W x D x H)
PMIB02	482.6 mm x 70 mm x 43.6 mm

19.12.3.3 Checking Environment Monitoring Cables

The SMU01B monitoring module monitors the environment parameters collected by the sensors in real time, and reports the alarm information to the control system. During the check, ensure that the cable connections from the sensors to the ports of the sensor transfer box, from the sensor transfer box to the monitoring module, and from the monitoring module to the control system are correct.

Checking Environment Monitoring Cables of the N66E-22 Cabinet

Figure 19-52 shows the connections of environment monitoring cables in the N66E-22 cabinet. Table 19-43 describes the cable connections. By referring to the connection diagram, check whether the connections from the sensors to the sensor transfer box (PMIB02), from the sensor transfer box to the monitoring module, and from the monitoring module to the control system are correct.

 Copyright © Huawei Technologies Co., Ltd.

Figure 19-52 Connections of environment monitoring cables in the N66E-22 cabinet

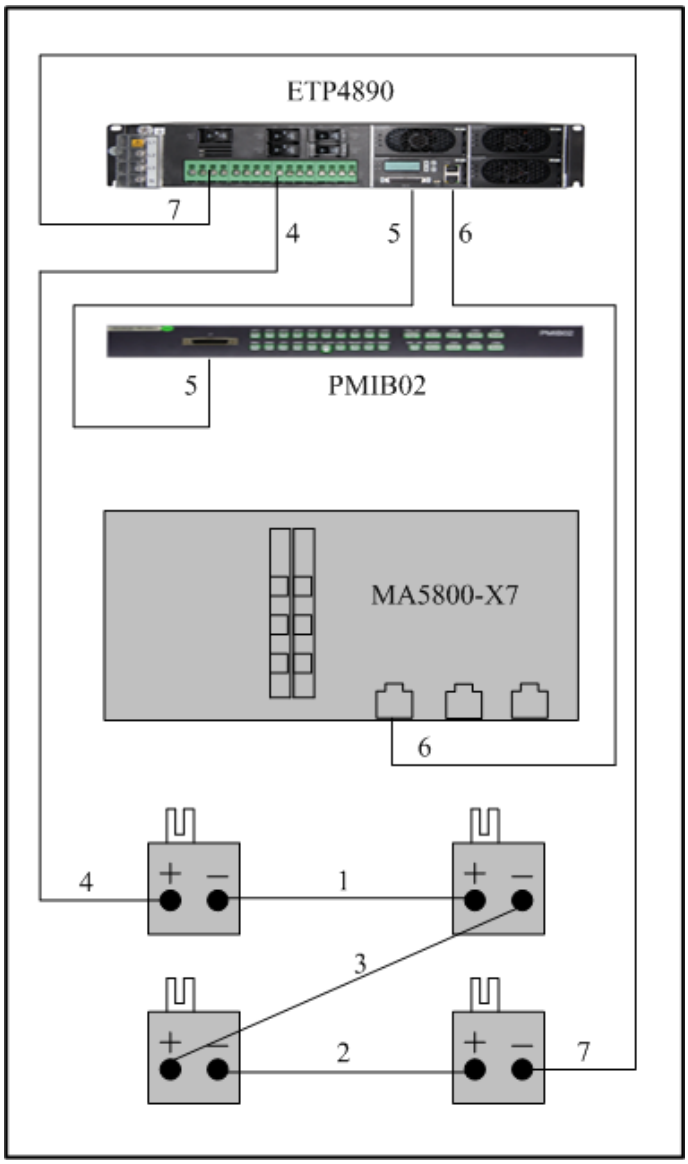


Table 19-43 Connections of environment monitoring cables in the N66E-22 cabinet

Cable	From	To
Monitoring cable of the ETP4890 to the control board	RS485/RS232 port or COM port on the ETP4890	ESC port on the control board

 Copyright © Huawei Technologies Co., Ltd.

Cable	From	To
Monitoring cable of the PMIB02 to the ETP4890	J1 port on the PMIB02	DB50 port on the ETP4890
Monitoring cable of the battery temperature sensor	VBTEM2 terminal of the PMIB02	Battery temperature sensor

19.12.3.4 Configuring the Environment Monitoring Parameters of the ETP4890

This topic describes how to configure the environment monitoring parameters of the ETP4890 through the CLI.

Mapping Between Monitoring Parameters and Device Ports

The ETP4890 monitoring scheme can be used in the N66E-22 cabinet. **Table 19-44** describes the mapping between the monitoring parameters displayed

NOTE

Before adding a user-defined analog or monitoring digital parameter, make sure that the port corresponding to this parameter is properly connected to an environment monitoring cable.

Table 19-44 Mapping between the monitoring parameters displayed on the host and the ports on the sensor transfer box

Monitoring Parameter Displayed on the Host	Device Port	Application in the N66E-22 Cabinet
Digital 0	JTD1	Not connected by default, used to add a user-defined monitoring digital parameter
Digital 1	JTD2	Not connected by default, used to add a user-defined monitoring digital parameter
Digital 2	JTD3	Not connected by default, used to add a user-defined monitoring digital parameter
Digital 4	JTD5	Not connected by default, used to add a user-defined monitoring digital parameter
Battery Tem	VBTEM2	Battery temperature sensor

 Copyright © Huawei Technologies Co., Ltd.

Procedure

Step 1 Log in to the device through the maintenance terminal and add an ETP4890 (SMU).

Run the **display emu** command to query the ID of the ETP4890.

Step 2 Check the default configurations of the power system according to the queried EMU ID.

Run the **interface emu** command to enter environment monitoring mode and then run the **display power system parameter** and **display power environment parameter** commands to query the default configurations.

Step 3 Configure parameters required for lead-acid batteries.

Run the **power battery parameter** command to configure parameters required for batteries, including the management parameters, charging parameters, power-off parameters, and high-temperature power-off parameters. You can use default values of these parameters for configuration.

Step 4 Run the **power battery temperature** command to configure the upper/lower temperature test limits, temperature compensation coefficient, and the upper/lower alarm temperature limits of the battery set.

Step 5 (Optional) Configure discharging testing parameters.

- Run the **power battery-test** command to configure the discharging testing parameters, including the automatic power-off test, scheduled automatic test, test interval for automatic discharging, test time for delayed discharging, voltage at the stop of the test, number of batteries in a battery set, time limit for the discharging test, time limit for simplified discharging test, alarm thresholds for discharging efficiencies. You can use default values of these parameters for configuration.
- Run the **display power battery-test info** command to query the discharge testing information about the batteries.

Step 6 Run the **power off** command to configure the battery set power-off voltage parameters

Step 7 Run the **power temperature-off** command to configure the battery high-temperature power-off parameters.

Step 8 Configure the environment monitoring parameters of the SMU system.

Run the **power environment** command to configure the upper alarm threshold and the upper and lower test thresholds for the temperatures and humidity. After these parameters are set, the system reports alarms if the working environment of the power system does not meet requirements.

Step 9 Configure required parameters for the rectifier module.

Run the **power module-num** command to configure the required parameters for the rectifier module.

Step 10 Configure the external extended digital variables of the power system.

 Copyright © Huawei Technologies Co., Ltd.

Run the **power outside-digital** command to configure the external extended digital variables of the power system.

Step 11 Query the information about the configuration parameters and environment parameters of the power system.

Run the **display power system parameter** and **display power environment parameter** commands to query the configuration parameters and environment parameters of the power system.

Step 12 Run the **display power alarm** command to query the alarm information, and ensure that no alarm is generated for the monitored objects except the door status.

Step 13 Save the data.

Run the **quit** command to exit the SMU mode, and run the **save** command to save the data.

----End

Result

After the configuration, the ETP4890 power system works properly and monitors the configured parameters. If the status of a monitored parameter becomes abnormal, an alarm is generated on the device.

Data Plan

Table 19-45 provides the data plan for configuring the monitoring parameters of the ETP4890.

Table 19-45 Data plan for configuring the power monitoring using the ETP4890

Item	Data Plan for the N66E-22 Cabinet
EMU	Type: SMU SN: 0 Subnode ID: 0 **NOTE** • The subnode ID must be the same as the subnode setting of the corresponding DIP switches on the EMU, but the subnode ID must be different from IDs of the other subnodes on the same bus. • The EMU ID is the same as that queried by running the **display emu** command.
Charging parameters of the battery	Charging mode of the battery: automatic Equalized charging voltage of the battery (V): 56.5 Float charging voltage of the battery (V): 53.5

 Copyright © Huawei Technologies Co., Ltd.

Item	Data Plan for the N66E-22 Cabinet
Battery management parameters	Current-limiting coefficient for battery charging: 0.15 Interval of battery equalized charging (day): 60 Number of battery sets: 1 Capacity of the battery set: ● Two 75 Ah battery sets: 150 Ah ● Two 100 Ah battery sets: 200 Ah
The upper/lower temperature test limits or temperature compensation coefficient of the battery set	Upper temperature threshold of the battery set (°C): 80 Lower temperature threshold of the battery set (°C): -20 Temperature compensation coefficient of the battery set (mV): 80
The upper/lower temperature alarm limits of the battery set	Upper temperature alarm limit of the battery set (°C): 50 Lower temperature alarm limit of the battery set (°C): -15
Battery discharging test parameters	Battery auto-discharging test: auto-test Battery set auto-discharging test permission status: permit Battery auto-discharging test period (day): 180
battery set power-off voltage parameters	Battery set power-off permission status: permit Battery set power-off voltage (V): 43
Battery high-temperature power-off parameters	Battery high-temperature power-off permission status: permit Temperature for battery high-temperature power-off (°C): 53
Environment temperature parameters	Upper alarm threshold of the temperature (°C): 68 Lower alarm threshold of the temperature (°C): -5 Upper test threshold of the temperature (°C): 80 Lower test threshold of the temperature (°C): -20
Environment humidity parameters	Upper alarm threshold of the humidity (% RH): 80 Lower alarm threshold of the humidity (% RH): 0 Upper test threshold of the humidity (% RH): 100 Lower test threshold of the humidity (% RH): 0
Power rectifier module parameters	The number of the power rectifier modules is configured according to the actual value. ETP4890 supports up to 3 rectifier modules.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

The preceding data is configured according to the actual requirements. When the actually planned value of a parameter is the same as the default value, it is not necessary to configure the parameter.

Example

Assume that the N66E-22 cabinet is configured with two 75 AH battery sets, the ETP4890 is configured with 3 rectifier modules. To configure the environment monitoring parameters of the ETP4890 after connected, do as follows:

```
huawei(config)#display emu
huawei(config)#interface emu 1
huawei(config-if-smu-1)#display power system parameter
huawei(config-if-smu-1)#display power environment parameter
huawei(config-if-smu-1)#power battery parameter 1 0.15 60 200
huawei(config-if-smu-1)#power battery temperature 80 -20 80
huawei(config-if-smu-1)#power battery temperature alarm_upper-limit 50 alarm_lower-limit -15
huawei(config-if-smu-1)#power battery-test auto-test permit 180
huawei(config-if-smu-1)#display power battery-test info 1
huawei(config-if-smu-1)#power off battery-off-state permit battery-off-voltage 43
huawei(config-if-smu-1)#power temperature-off battery-off-state permit battery-off-temperature 53
huawei(config-if-smu-1)#power environment temperature 68 -5 80 -20
  This command is invalid unless in the contition of install the sensor, would
you continue? (y/n)[n]:y
huawei(config-if-smu-1)#power environment humidity 80 0 100 0
  This command is invalid unless in the contition of install the sensor, would
you continue? (y/n)[n]:y
huawei(config-if-smu-1)#power module-num 3 1 2 3

huawei(config-if-smu-1)#display power system parameter
huawei(config-if-smu-1)#display power environment parameter
huawei(config-if-smu-1)#display power alarm
huawei(config-if-smu-1)#quit
huawei(config)#save
```

19.12.4 ETP48200 Monitoring Solution

The ETP48200 monitoring solution is implemented by the SMU11C monitoring module in the ETP48200 power system. In this solution, the SMU11C monitoring module reports the monitored parameters reported by the sensor transfer box to the control system.

19.12.4.1 ETP48200-B2A1 Power System

The ETP48200-B2A1 (abbreviated as ETP48200) is an embedded power system that supplies power to –48 V DC telecom equipment. It has a maximum output current of 150 A.

Function

The ETP48200-B2A1 power system supports the following functions:

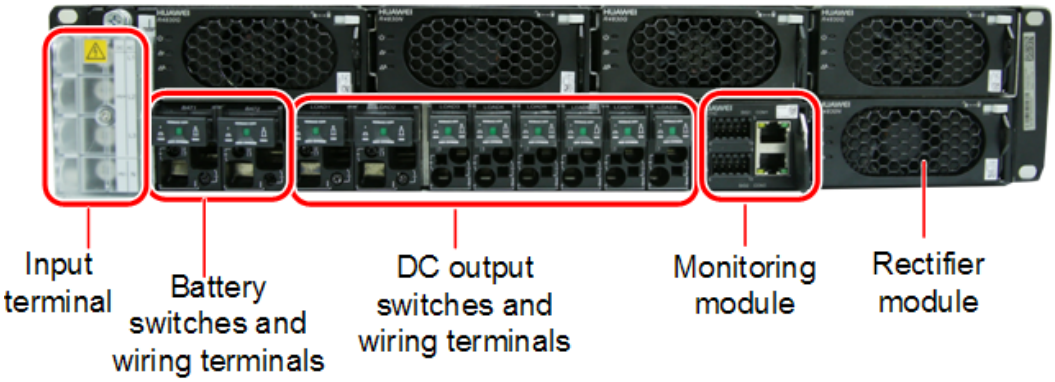
- Supports AC or HVDC input. AC input voltage range: 85–300 V AC; HVDC input voltage range: 85–420 V DC.
- Comprehensive battery management.
- Communication and alarm functions; remote management and online upgrade.

 Copyright © Huawei Technologies Co., Ltd.

- One CAN port, RS485, or FE port for network connections.
- Hot-swappable rectifiers and monitoring module.

Appearance

Figure 19-53 Appearance of ETP48200-B2A1power system



Configuration

Table 19-46 Component configuration of the ETP48200-B2A1power system

Component	Configuration
Subrack	2 U power system
Power distribution unit (PDU)	• AC or DC input: M6 OT wiring terminal • DC output: supports a maximum of nine 63 A or lower circuit breakers or four 80 A and three 63 A circuit breakers. • Battery route: Two 100 A circuit breakers are supported.
Rectifier	A maximum of five R4830G1s, or R4815G1s can be configured.
Monitoring module	SMU11C

The ETP48200-B2A1 power system can be configured with 5 rectifiers, which are connected in parallel for output.

Table 19-47 shows the mapping between the quantity of power supply modules in the ETP48200-B2A1 power system and the maximum output current.

 Copyright © Huawei Technologies Co., Ltd.

Table 19-47 Mapping between the quantity of power supply modules of the ETP48200-B2A1 power system and the maximum output current

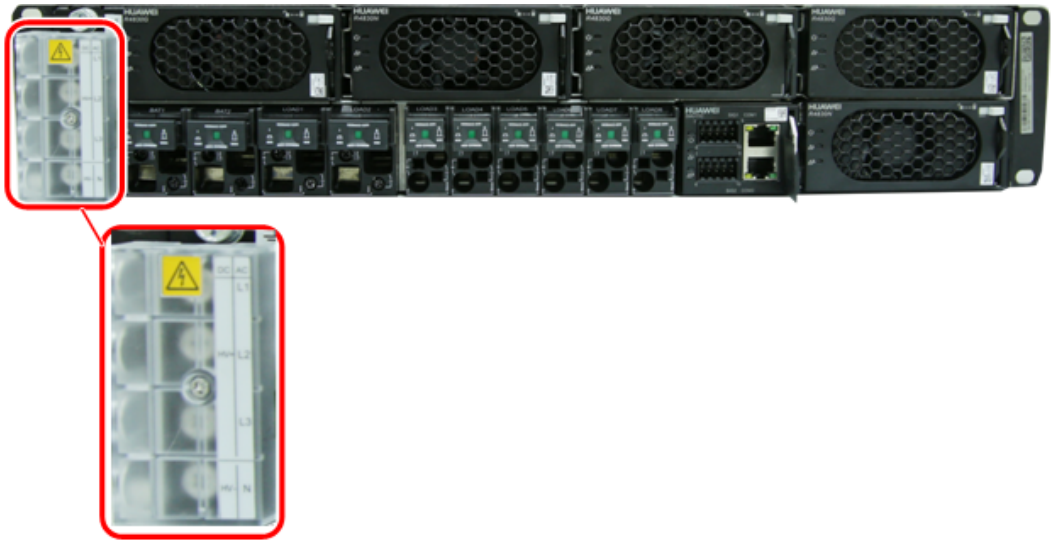
Quantity of Power Supply Modules	Maximum Output Current
1	30 A
2	60 A
3	90 A
4	120 A
5	150 A

Input Terminal

The ETP48200-B2A21power system supports single-phase, three-phase, four-wire, dual-live-wire, or HVDC input.

Figure 19-54 shows the input terminals in the ETP48200-B2A1 power system.

Figure 19-54 Input terminals in the ETP48200-B2A1 power system



Output Terminal

The ETP48200-B2A1 power system supports flexible configuration and supports 63A or lower current. 20 mm width; 80A, 100A, and 125A: 30 mm width; The 20 mm and 30 mm cover can be flexibly interchanged. Two 30 mm circuit breakers can be replaced with three 20mm shells.

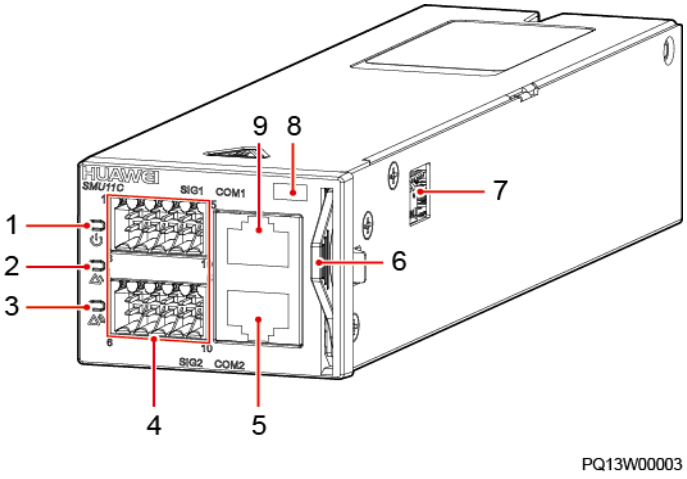
Figure 19-55 shows the output terminals in the ETP48200-B2A1 power system.

 Copyright © Huawei Technologies Co., Ltd.

Figure 19-55 Output terminals in the ETP48200-B2A1 power system

Monitoring Module SMU11C

Figure 19-56 SMU11C appearance (without terminals)



(1) Running indicator	(2) Minor alarm indicator	(3) Major alarm indicator
(4) Wiring terminals	(5) Communications port COM2	(6) Handle
(7) Dual-in-line package (DIP) switch	(8) Position of the SN code	(9) Communications port COM1

Table 19-48 Description of the indicators on the SMU11C panel

Name	Color	Status	Description
Running indicator	Green	Off	The SMU is faulty or has no power input.

 Copyright © Huawei Technologies Co., Ltd.

Name	Color	Status	Description
		Blinking slowly (0.5 Hz)	The SMU is running and communicating with the host properly.
		Blinking fast (4 Hz)	The SMU is running properly but fails to communicate with the host properly.
Minor alarm indicator	Yellow	Off	No minor alarm or warning is generated.
		Steady on	A minor alarm or warning is generated.
Major alarm indicator	Red	Off	No critical or major alarm is generated.
		Steady on	A critical or major alarm is generated.

Figure 19-57 Wiring terminals of monitoring module SMU11C

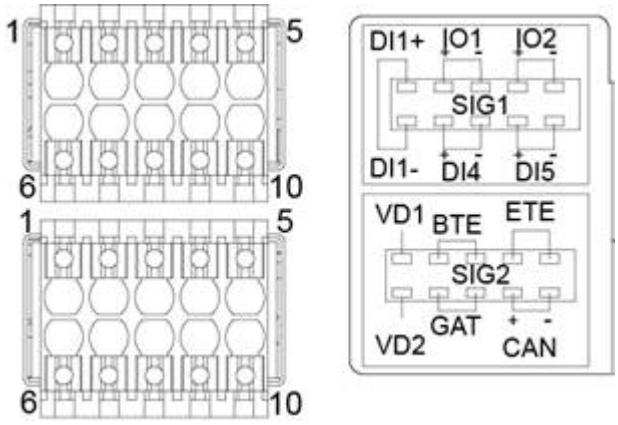


Table 19-49 Pin definitions for SIG1 wiring terminals of monitoring module SMU11C

Pin	Signal	Description
1	DI1+	Dry contact input
6	DI1-	
7	DI4+	
8	DI4-	
9	DI5+	
10	DI5-	

 Copyright © Huawei Technologies Co., Ltd.

Pin	Signal	Description
2	IO1+	Dry contact input/Dry contact output (When used as a dry contact input, the alarm condition is as follows: normal when open, alarm when closed. When used as a dry contact output, the alarm action is as follows: open when normal, closed when alarm.)
3	IO1-	
4	IO2+	
5	IO2-	

Table 19-50 Pin definitions for SIG2 wiring terminals of monitoring module SMU11C

Pin	Signal	Description
1	VD1	Midpoint voltage detection port 1
6	VD2	Midpoint voltage detection port 2
2	BTE	Battery temperature sensor port
3		
4	ETE	Ambient temperature sensor port
5		
7	GAT	Door status sensor port
8		
9	CAN+	CAN communications port
10	CAN-	

Table 19-51 Communications port description of monitoring module SMU11C

Communications Port	Communications Parameter	Communications Protocol	Function
COM1	Baud rate: 9600 bit/s, 19200 bit/s, or 115200 bit/s	Master/slave protocols	Connects to an upper-level network management system (NMS).

 Copyright © Huawei Technologies Co., Ltd.

Communications Port	Communications Parameter	Communications Protocol	Function
COM2	Baud rate: 9600 bit/s, 19200 bit/s, or 115200 bit/s	Master/slave and Modbus protocols	Connects to an upper-level NMS or intelligent equipment.
NOTE All these ports are protected by a security mechanism.			

Figure 19-58 COM port pins on the SMU11C panel

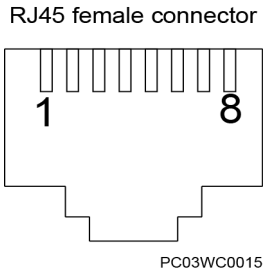


Table 19-52 Pin definitions for the COM1 port of monitoring module SMU11C

Pin	Signal	Description
1	TX+	Transmits data over RS485.
2	TX-	
3	RX232	Receives data over RS232.
4	RX+	Receives data over RS485.
5	RX-	
6	PGND	Grounds the port (PE).
7	TX232	Transmits data over RS232.
8	Reserved	-

Table 19-53 Pin definitions for the COM2 port of monitoring module SMU11C

Pin	Signal	Description
1	TX+	Transmits data over RS485.
2	TX-	
3	12 V	12 V output (rated current: 500 mA)

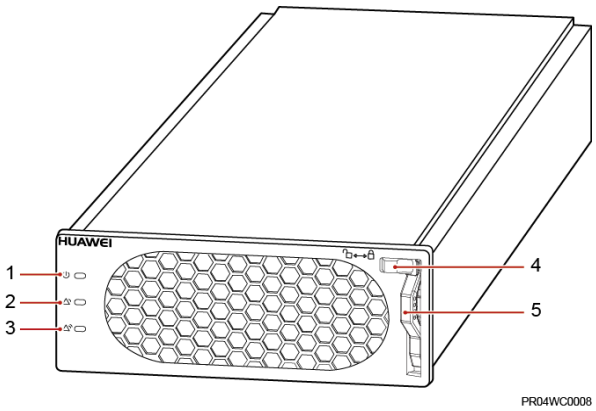
 Copyright © Huawei Technologies Co., Ltd.

Pin	Signal	Description
4	RX+	Receives data over RS485.
5	RX-	
6	I^2C_SCL	I^2C clock signal
7	I^2C_SDA	I^2C data signal
8	PGND	Grounds the port (PE).

Rectifier Module

A rectifier module converts AC or DC input power into stable DC power.

Figure 19-59 Rectifier module appearance



(1) Run indicator	(2) Alarm indicator	(3) Fault indicator
(4) Locking latch	(5) Handle	

Table 19-54 Rectifier module indicator description

Indicator	Color	Status	Description
Run indicator	Green	Steady on	The rectifier has an AC power input.
		Off	The rectifier has no AC power input.
			The rectifier is faulty.
		Blinking at 0.5 Hz	The rectifier is being queried.
		Blinking at 4 Hz	The rectifier is loading an application program.

 Copyright © Huawei Technologies Co., Ltd.

Indicator	Color	Status	Description
Alarm indicator	Yellow	Off	No alarm has been generated.
		Steady on	• The rectifier has generated an alarm due to ambient overtemperature. • The rectifier has generated an alarm for shutdown due to ambient overtemperature or undertemperature.
			The rectifier is protecting itself against AC input overvoltage or undervoltage.
			The rectifier is hibernating.
		Blinking at 0.5 Hz	The communication between the rectifier and the SMU has been interrupted.
Fault indicator	Red	Off	The rectifier is running properly.
		Steady on	The rectifier has been locked out due to output overvoltage.
			The rectifier has no output due to an internal fault.

Technical Specifications

The following table shows the technical specifications of the ETP48200-B2A1power system.

Table 19-55 Technical specifications

Category	Item	Specifications
Environmental condition	Operating temperature	–40°C to +65°C; when the operating temperature ranges from 55°C to 65°C, rectifier power is derated automatically.
	Transportation/ Storage temperature	–40°C to +70°C
	Operating humidity	5%–90% RH (non-condensing)
	Storage humidity	5%–95% RH (non-condensing)

 Copyright © Huawei Technologies Co., Ltd.

Category	Item	Specifications
	Altitude	0–4000 m When the altitude ranges from 2000 m to 4000 m, high temperature derating applies and the operating temperature decreases by 1°C for each additional 200 m.
AC input	Input system	220/380 V AC three-phase four-wire (compatible with 220 V AC single-phase and 110 V AC dual-live wire)
	Rated voltage	200/346–240/415V AC
	Voltage range	85/148–300/520V AC
	Input frequency	45–65 Hz (rated frequency: 50 Hz/60 Hz)
DC input	Input system	HVDC (HV+, HV–) input
	Rated voltage	380 V DC
	Voltage range	85–420 V DC
DC output	Output voltage range	–42 V DC to –58 V DC
	Maximum output power	10000 W
	Regulated voltage precision	≤ ±1%
	Peak-to-peak noise voltage	≤ 200 mV
AC input protection	AC input over-voltage protection threshold	> 300 V AC
	AC input overvoltage recovery threshold	When the voltage is restored to 290 V AC, the output resumes.
	AC input under-voltage protection threshold	< 80 V AC
	AC input under-voltage recovery threshold	When the voltage is restored to 85 V AC, the output resumes.
DC input protection	DC input overvoltage protection threshold	> 420 V DC

 Copyright © Huawei Technologies Co., Ltd.

Category	Item	Specifications
	DC input overvoltage recovery threshold	When the voltage is restored to 414 V DC, the output resumes.
DC output protection	DC output overvoltage protection threshold	Range: 56–60 V DC
Structure	Dimensions (H x W x D)	86.1 mm x 442 mm x 255 mm (without mounting ears)
	Weight	≤ 16 kg (including five rectifiers and one monitoring module)
	Protection level	IP20
	Installation mode	Installed horizontally or vertically in a cabinet
	Maintenance mode	Maintained from the front
	Cooling	Free cooling

19.12.4.2 Checking the Environment Monitoring Cable

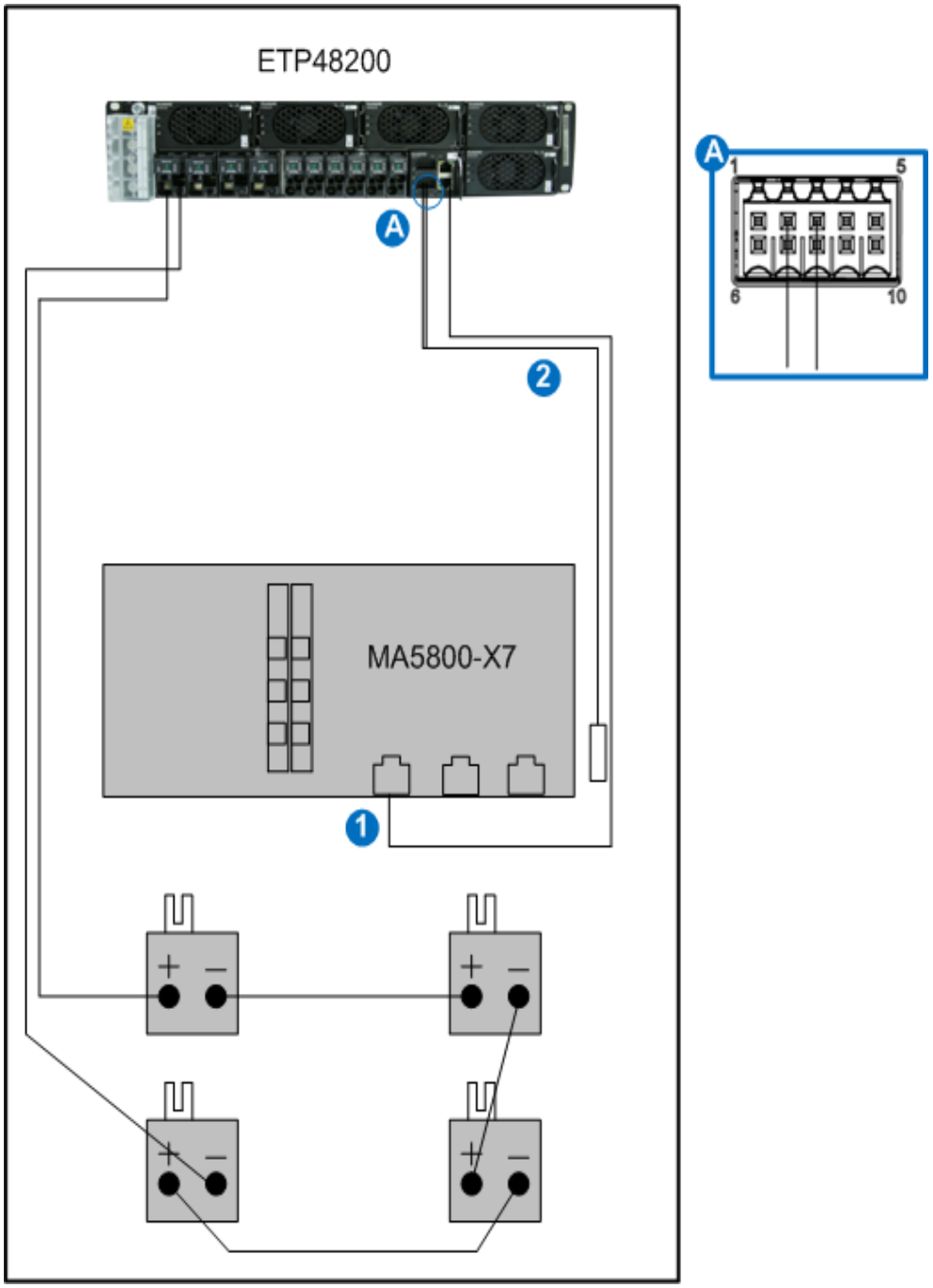
The SMU11C monitoring module monitors environment monitoring parameters collected by sensors in real time, and reports alarms to the control system. During the check, ensure that the cable connections from the sensors to the ports of the sensor transfer box, from the sensor transfer box to the monitoring module, and from the monitoring module to the control system are correct.

Checking Environment Monitoring Cables of the N66E-22 Cabinet

Figure 19-60 shows the connections of the N66E-22 environment monitoring cables, and **Table 19-56** describes the cable connections. By referring to the connection diagram, check whether the connections from the sensors to the sensor transfer box (PMIB02), from the sensor transfer box to the monitoring module, and from the monitoring module to the control system are correct.

 Copyright © Huawei Technologies Co., Ltd.

Figure 19-60 Connections of environment monitoring cables in the N66E-22 cabinet



 Copyright © Huawei Technologies Co., Ltd.

Table 19-56 Connections of environment monitoring cables in the N66E-22 cabinet

Cable Name	From	To
Monitoring cable between the ETP48200 and the control board	COM2 port on the ETP48200 monitoring module	ESC port on the control board
Monitoring cable of the battery temperature sensor	ETE port on the ETP48200 monitoring module	Battery temperature sensor

19.12.4.3 Configuring the Environment Monitoring Parameters of the ETP48200

This topic describes how to configure the environment monitoring parameters of the ETP4890 through the CLI.

Mapping Between Monitoring Parameters and Device Ports

The ETP4890 monitoring scheme can be used in the N66E-22 cabinet. **Table 19-57** describes the mapping between the monitoring parameters displayed

NOTE

Before adding a user-defined analog or monitoring digital parameter, make sure that the port corresponding to this parameter is properly connected to an environment monitoring cable.

Table 19-57 Mapping between the monitoring parameters displayed on the host and the ports on the sensor transfer box

Monitoring Parameter Displayed on the Host	Device Port	Application in the N66E-22 Cabinet
Digital 0	JTD1	Not connected by default, used to add a user-defined monitoring digital parameter
Digital 1	JTD2	Not connected by default, used to add a user-defined monitoring digital parameter
Digital 2	JTD3	Not connected by default, used to add a user-defined monitoring digital parameter

 Copyright © Huawei Technologies Co., Ltd.

Monitoring Parameter Displayed on the Host	Device Port	Application in the N66E-22 Cabinet
Digital 4	JTD5	Not connected by default, used to add a user-defined monitoring digital parameter
Battery Tem	VBTEM2	Battery temperature sensor

Procedure

Step 1 Log in to the device through the maintenance terminal and add an ETP4890 (SMU).

Run the **display emu** command to query the ID of the ETP4890.

Step 2 Check the default configurations of the power system according to the queried EMU ID.

Run the **interface emu** command to enter environment monitoring mode and then run the **display power system parameter** and **display power environment parameter** commands to query the default configurations.

Step 3 Configure parameters required for lead-acid batteries.

Run the **power battery parameter** command to configure parameters required for batteries, including the management parameters, charging parameters, power-off parameters, and high-temperature power-off parameters. You can use default values of these parameters for configuration.

Step 4 Run the **power battery temperature** command to configure the upper/lower temperature test limits, temperature compensation coefficient, and the upper/lower alarm temperature limits of the battery set.

Step 5 (Optional) Configure discharging testing parameters.

- Run the **power battery-test** command to configure the discharging testing parameters, including the automatic power-off test, scheduled automatic test, test interval for automatic discharging, test time for delayed discharging, voltage at the stop of the test, number of batteries in a battery set, time limit for the discharging test, time limit for simplified discharging test, alarm thresholds for discharging efficiencies. You can use default values of these parameters for configuration.
- Run the **display power battery-test info** command to query the discharge testing information about the batteries.

Step 6 Run the **power off** command to configure the battery set power-off voltage parameters

Step 7 Run the **power temperature-off** command to configure the battery high-temperature power-off parameters.

Step 8 Configure the environment monitoring parameters of the SMU system.

Run the **power environment** command to configure the upper alarm threshold and the upper and lower test thresholds for the temperatures and humidity. After

 Copyright © Huawei Technologies Co., Ltd.

these parameters are set, the system reports alarms if the working environment of the power system does not meet requirements.

Step 9 Configure required parameters for the rectifier module.

Run the **power module-num** command to configure the required parameters for the rectifier module.

Step 10 Configure the external extended digital variables of the power system.

Run the **power outside-digital** command to configure the external extended digital variables of the power system.

Step 11 Query the information about the configuration parameters and environment parameters of the power system.

Run the **display power system parameter** and **display power environment parameter** commands to query the configuration parameters and environment parameters of the power system.

Step 12 Run the **display power alarm** command to query the alarm information, and ensure that no alarm is generated for the monitored objects except the door status.

Step 13 Save the data.

Run the **quit** command to exit the SMU mode, and run the **save** command to save the data.

----End

Result

After the configuration, the ETP4890 power system works properly and monitors the configured parameters. If the status of a monitored parameter becomes abnormal, an alarm is generated on the device.

Data Plan

Table 19-58 provides the data plan for configuring the monitoring parameters of the ETP4890.

Table 19-58 Data plan for configuring the power monitoring using the ETP4890

Item	Data Plan for the N66E-22 Cabinet
EMU	Type: SMU SN: 0 Subnode ID: 0 **NOTE** • The subnode ID must be the same as the subnode setting of the corresponding DIP switches on the EMU, but the subnode ID must be different from IDs of the other subnodes on the same bus. • The EMU ID is the same as that queried by running the **display emu** command.

 Copyright © Huawei Technologies Co., Ltd.

Item	Data Plan for the N66E-22 Cabinet
Charging parameters of the battery	Charging mode of the battery: automatic Equalized charging voltage of the battery (V): 56.5 Float charging voltage of the battery (V): 53.5
Battery management parameters	Current-limiting coefficient for battery charging: 0.15 Interval of battery equalized charging (day): 60 Number of battery sets: 1 Capacity of the battery set: ● Two 75 Ah battery sets: 150 Ah ● Two 100 Ah battery sets: 200 Ah
The upper/lower temperature test limits or temperature compensation coefficient of the battery set	Upper temperature threshold of the battery set (°C): 80 Lower temperature threshold of the battery set (°C): -20 Temperature compensation coefficient of the battery set (mV): 80
The upper/lower temperature alarm limits of the battery set	Upper temperature alarm limit of the battery set (°C): 50 Lower temperature alarm limit of the battery set (°C): -15
Battery discharging test parameters	Battery auto-discharging test: auto-test Battery set auto-discharging test permission status: permit Battery auto-discharging test period (day): 180
battery set power-off voltage parameters	Battery set power-off permission status: permit Battery set power-off voltage (V): 43
Battery high-temperature power-off parameters	Battery high-temperature power-off permission status: permit Temperature for battery high-temperature power-off (°C): 53
Environment temperature parameters	Upper alarm threshold of the temperature (°C): 68 Lower alarm threshold of the temperature (°C): -5 Upper test threshold of the temperature (°C): 80 Lower test threshold of the temperature (°C): -20
Environment humidity parameters	Upper alarm threshold of the humidity (% RH): 80 Lower alarm threshold of the humidity (% RH): 0 Upper test threshold of the humidity (% RH): 100 Lower test threshold of the humidity (% RH): 0

 Copyright © Huawei Technologies Co., Ltd.

Item	Data Plan for the N66E-22 Cabinet
Power rectifier module parameters	The number of the power rectifier modules is configured according to the actual value. ETP4890 supports up to 3 rectifier modules.

NOTE

The preceding data is configured according to the actual requirements. When the actually planned value of a parameter is the same as the default value, it is not necessary to configure the parameter.

Example

Assume that the N66E-22 cabinet is configured with two 75 AH battery sets, the ETP4890 is configured with 3 rectifier modules. To configure the environment monitoring parameters of the ETP4890 after connected, do as follows:

```
huawei(config)#display emu
huawei(config)#interface emu 1
huawei(config-if-smu-1)#display power system parameter
huawei(config-if-smu-1)#display power environment parameter
huawei(config-if-smu-1)#power battery parameter 1 0.15 60 200
huawei(config-if-smu-1)#power battery temperature 80 -20 80
huawei(config-if-smu-1)#power battery temperature alarm_upper-limit 50 alarm_lower-limit -15
huawei(config-if-smu-1)#power battery-test auto-test permit 180
huawei(config-if-smu-1)#display power battery-test info 1
huawei(config-if-smu-1)#power off battery-off-state permit battery-off-voltage 43
huawei(config-if-smu-1)#power temperature-off battery-off-state permit battery-off-temperature 53
huawei(config-if-smu-1)#power environment temperature 68 -5 80 -20
  This command is invalid unless in the contition of install the sensor, would
you continue? (y/n)[n]:y
huawei(config-if-smu-1)#power environment humidity 80 0 100 0
  This command is invalid unless in the contition of install the sensor, would
you continue? (y/n)[n]:y
huawei(config-if-smu-1)#power module-num 3 1 2 3

huawei(config-if-smu-1)#display power system parameter
huawei(config-if-smu-1)#display power environment parameter
huawei(config-if-smu-1)#display power alarm
huawei(config-if-smu-1)#quit
huawei(config)#save
```

19.12.5 Fan Tray Monitoring Solution

Fans are monitored through the fan monitoring board that reports the fan status to the control system.

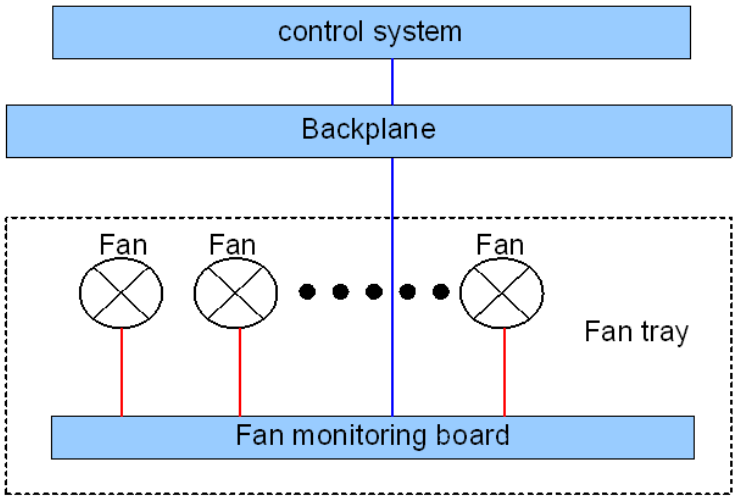
19.12.5.1 Fan Tray Monitoring Principles

This topic describes the principles of monitoring the fan tray through the fan monitoring board.

Figure 19-61 shows the fan tray monitoring principles.

 Copyright © Huawei Technologies Co., Ltd.

Figure 19-61 Fan tray monitoring principles



NOTE

- The fans connect to the fan monitoring board through the signal cable. The fan monitoring board monitors the working status of the fans.
- The fan tray is connected to the backplane of the shelf through cables directly or indirectly.
- The fan monitoring board reports the collected status to the control system through the backplane.

19.12.5.2 MA5800-X17 Fan Tray

The fan tray of the MA5800-X17 service subrack has eight fans and fan monitoring board FMLA, and functions heat dissipation, monitoring, and fan speed adjustment, which ensures that the device works at a stable temperature.

Specification

Item	Specification
BOM number	02122392
Dimensions (W x D x H)	503 mm x 284 mm x 76 mm
Weight	4.40 kg (Fan tray)

 Copyright © Huawei Technologies Co., Ltd.

Item	Specification
Power consumption **NOTE** The power consumption of the fan board is measured under the following conditions: • **Static power consumption**: power consumption when the fan speed is 0. • **Maximum power consumption**: power consumption of the fan board and fan assembly when the fan speed is 100%.	• Static: 20 W • Maximum: 139 W

Function

The functions of the fan tray are as follows:

- **Heat dissipation**

  The fan tray is at the top of the service subrack and exhausts hot air for heat dissipation. The cool air flows to the subrack from the bottom of the subrack and then is exhausted from the top of the subrack after passing the boards.
- **Monitoring**

  The fan tray is configured with the fan monitoring board to detect whether the fans are working in the normal state. The fan monitoring board also provides the port for communication with the control board. The detected information is transmitted to the control board through the fan monitoring board periodically.
- **Speed adjustment**

  The rotating speed of the fans can be adjusted according to the detected temperature automatically or be adjusted by setting the related data manually.

 Copyright © Huawei Technologies Co., Ltd.

Indicator

Indicator	Color	Status	Meaning	Operation Description
STATUS	Yellow	Blinking quickly (on for 0.125 s and off for 0.125 s repeatedly)	The fan tray is not registered, is being loaded, or fails to communicate with the control board.	• If the fan tray is not registered or is being loaded, no action is required. • If the fan tray fails to communicate with the control board, check the communication connection between the fan tray and the control board.
	Green	Blinking	The fan tray works in the normal state.	No action is required.
	Yellow	Blinking slowly (on for 1 s and off for 1 s repeatedly)	The fan tray generates alarms that do not affect services.	Handle it based on the corresponding alarm.
	Red	Blinking	The fan tray is faulty or the fan tray generates an over-temperature alarm.	• Replace the faulty fan tray. • Increase the fan speed to lower the temperature.

Fan Speed Adjustment

Set the fan speed adjustment mode to "automatic" or "manual" in the command line interface (CLI). After the fans are installed, they work in automatic mode by default.

- Automatic mode
  - Control-system-triggering: The control system automatically adjusts the fan speed according to the board temperatures for energy conservation.

 Copyright © Huawei Technologies Co., Ltd.

– Monitoring-board-triggering: The control system adjusts the fan speed according to the temperature information collected by the monitoring board.

NOTE

The automatic mode has two variants (automatically selected by the system): control-system-triggering and monitoring-board-triggering.

- Manual mode

  Commands are executed to adjust the fan speed. The levels range from 0 to 6, level 0 being the lowest speed and level 6 being the highest speed. Alternatively, the fan speed can be set from 20% to 100%.

19.12.5.3 MA5800-X15 Fan Tray

The fan tray of the MA5800-X15 service subrack has six fans or eight fans and fan monitoring board FMLA, and functions heat dissipation, monitoring, and fan speed adjustment, which ensures that the device works at a stable temperature.

Figure 19-62 Appearance

Specification

Item	Specification
BOM number	02122393 (Configured with 6 fans) 02122797 (Configured with 8 fans)
Dimensions (W x D x H)	448 mm x 284 mm x 76 mm
Weight	About 4.85kg (Configured with 6 fans) About 4.84kg (Configured with 8 fans)

 Copyright © Huawei Technologies Co., Ltd.

Item	Specification
Power consumption **NOTE** The power consumption of the fan board is measured under the following conditions: • **Static power consumption**: power consumption when the fan speed is 0. • **Maximum power consumption**: power consumption of the fan board and fan assembly when the fan speed is 100%.	Configured with 6 fans: • Static: 18W • Maximum: 225W Configured with 8 fans: • Static: 28W • Maximum: 354W

Function

The functions of the fan tray are as follows:

- **Heat dissipation**

  The fan tray is at the top of the service subrack and exhausts hot air for heat dissipation. The cool air flows to the subrack from the bottom of the subrack and then is exhausted from the top of the subrack after passing the boards.
- **Monitoring**

  The fan tray is configured with the fan monitoring board to detect whether the fans are working in the normal state. The fan monitoring board also provides the port for communication with the control board. The detected information is transmitted to the control board through the fan monitoring board periodically.
- **Speed adjustment**

  The rotating speed of the fans can be adjusted according to the detected temperature automatically or be adjusted by setting the related data manually.

 Copyright © Huawei Technologies Co., Ltd.

Indicator

Indicator	Color	Status	Meaning	Operation Description
STATUS	Yellow	Blinking quickly (on for 0.125 s and off for 0.125 s repeatedly)	The fan tray is not registered, is being loaded, or fails to communicate with the control board.	• If the fan tray is not registered or is being loaded, no action is required. • If the fan tray fails to communicate with the control board, check the communication connection between the fan tray and the control board.
	Green	Blinking	The fan tray works in the normal state.	No action is required.
	Yellow	Blinking slowly (on for 1 s and off for 1 s repeatedly)	The fan tray generates alarms that do not affect services.	Handle it based on the corresponding alarm.
	Red	Blinking	The fan tray is faulty or the fan tray generates an over-temperature alarm.	• Replace the faulty fan tray. • Increase the fan speed to lower the temperature.

Fan Speed Adjustment

Set the fan speed adjustment mode to "automatic" or "manual" in the command line interface (CLI). After the fans are installed, they work in automatic mode by default.

- Automatic mode
  - Control-system-triggering: The control system automatically adjusts the fan speed according to the board temperatures for energy conservation.

 Copyright © Huawei Technologies Co., Ltd.

  - Monitoring-board-triggering: The control system adjusts the fan speed according to the temperature information collected by the monitoring board.

    NOTE

    The automatic mode has two variants (automatically selected by the system): control-system-triggering and monitoring-board-triggering.

- Manual mode

  Commands are executed to adjust the fan speed. The levels range from 0 to 6, level 0 being the lowest speed and level 6 being the highest speed. Alternatively, the fan speed can be set from 30% to 100%.

19.12.5.4 MA5800-X7 Fan Tray

The fan tray of the MA5800-X7 service subrack has 6 fans and a fan monitoring board FMMA and supports heat dissipation, monitoring, and fan speed adjustment to ensure that the device works at a proper temperature.

Specification

Item	Specification
BOM number	02122461
Dimensions (H x W x D)	43.6 mm x 257.5 mm x 260.4 mm
Weight	2.00 kg (Fan tray)

 Copyright © Huawei Technologies Co., Ltd.

Item	Specification
Power consumption **NOTE** The power consumption of the fan board is measured under the following conditions: • **Static power consumption**: power consumption when the fan speed is 0. • **Maximum power consumption**: power consumption of the fan board and fan assembly when the fan speed is 100%.	• Static: 14 W • Maximum: 189 W

Function

The functions of the fan tray are as follows:

- **Heat dissipation**

  The fan tray is on the right side of the service subrack to exhaust hot air. The air flow is from left to right.

- **Monitoring**

  The fan monitoring board detects whether the fans are working in a normal state and sends the results to the control board periodically.

- **Speed adjustment**

  The rotating speed of the fans can be adjusted manually or automatically based on the detected temperature.

Indicator

Indicator	Color	Status	Meaning	Operation Description
STATUS	Yellow	Blinking quickly (on for 0.125 s and off for 0.125 s repeatedly)	The fan tray is not registered, is being loaded, or fails to communicate with the control board.	• If the fan tray is not registered or is being loaded, no action is required. • If the fan tray fails to communicate with the control board, check the communication connection between the fan tray and the control board.

 Copyright © Huawei Technologies Co., Ltd.

Indicator	Color	Status	Meaning	Operation Description
	Green	Blinking	The fan tray works in the normal state.	No action is required.
	Yellow	Blinking slowly (on for 1 s and off for 1 s repeatedly)	The fan tray generates alarms that do not affect services.	Handle it based on the corresponding alarm.
	Red	Blinking	The fan tray is faulty or the fan tray generates an over-temperature alarm.	• Replace the faulty fan tray. • Increase the fan speed to lower the temperature.

Fan Speed Adjustment

The fan speed adjustment mode can be set to "automatic" or "manual" in the command line interface (CLI). The default mode is "automatic".

NOTE

The fan tray is used for heat dissipation of the system so that the device can work at a normal temperature. When the ambient temperature exceeds the upper limit, the system automatically reports a high temperature alarm.

- Automatic mode
  - Based on the control system: The control system automatically adjusts the fan speed according to the board temperatures for energy conservation.
  - Based on the monitoring board: The control system adjusts the fan speed based on the temperature information collected by the monitoring board.

  NOTE

  When the automatic speed adjustment mode is used, the device automatically switches between adjustment based on the control system and the adjustment based on the monitoring board.
- Manual mode

 Copyright © Huawei Technologies Co., Ltd.

You can also manually adjust the fan speed to a level between level 0 (the lowest speed) and level 6 (the highest speed) or a percentage between 20% and 100% using commands.

19.12.5.5 MA5800-X2 Fan Tray

The fan tray of the MA5800-X2 service subrack has 2 fans and a fan monitoring board (FMSA) and supports heat dissipation, monitoring, and fan speed adjustment to ensure that the device works at a proper temperature.

Specification

Item	Specification
Dimensions (H x W x D)	43.6 mm x 286.7 mm x 88.1 mm
Weight	0.70 kg (Fan tray)
Power consumption **NOTE** The power consumption of the fan board is measured under the following conditions: • **Static power consumption**: power consumption when the fan speed is 0. • **Maximum power consumption**: power consumption of the fan board and fan assembly when the fan speed is 100%.	• Static: 6 W • Maximum: 63 W

Function

The functions of the fan tray are as follows:

- **Heat dissipation**

  The fan tray is on the right side of the service subrack to exhaust hot air. The air flow is from left to right.

- **Monitoring**

  The fan monitoring board detects whether the fans are working in a normal state and sends the results to the control board periodically.

- **Speed adjustment**

 Copyright © Huawei Technologies Co., Ltd.

The rotating speed of the fans can be adjusted manually or automatically based on the detected temperature.

Indicator

Indicator	Color	Status	Meaning	Operation Description
STATUS	Yellow	Blinking quickly (on for 0.125 s and off for 0.125 s repeatedly)	The fan tray is not registered, is being loaded, or fails to communicate with the control board.	• If the fan tray is not registered or is being loaded, no action is required. • If the fan tray fails to communicate with the control board, check the communication connection between the fan tray and the control board.
	Green	Blinking	The fan tray works in the normal state.	No action is required.
	Yellow	Blinking slowly (on for 1 s and off for 1 s repeatedly)	The fan tray generates alarms that do not affect services.	Handle it based on the corresponding alarm.
	Red	Blinking	The fan tray is faulty or the fan tray generates an over-temperature alarm.	• Replace the faulty fan tray. • Increase the fan speed to lower the temperature.

Fan Speed Adjustment

The fan speed adjustment mode can be set to "automatic" or "manual" in the command line interface (CLI). The default mode is "automatic".

- Automatic mode

 Copyright © Huawei Technologies Co., Ltd.

  - Based on the control system: The control system automatically adjusts the fan speed according to the board temperatures for energy conservation.
  - Based on the monitoring board: The control system adjusts the fan speed based on the temperature information collected by the monitoring board.
- Manual mode

  You can also manually adjust the fan speed to a level between level 0 (the lowest speed) and level 6 (the highest speed) or a percentage between 20% and 100% using commands.

19.12.5.6 Commissioning the Fan EMU

This topic describes how to commission the fan EMU to ensure that it monitors the environmental conditions of the fans of the device according to the actual conditions.

Context

NOTE

When the device is delivered, the fan EMU is correctly connected to the shelf. The connection need not be changed for the device commissioning. In certain cases, if the EMU needs to be configured in other shelves, reconnect the EMU. For details, see this topic.

The fan EMU is used to monitor the running status of the fans and to set the fan rotation speed according to actual conditions to ensure the normal heat dissipation of the device.

When commissioning the fan EMU, pay attention to the following points:

- The EMU sub-nodes are numbered from 0 to 31.
- When the system is configured with multiple EMUs simultaneously, make sure that the sub-nodes do not conflict with each other.
- It is recommended that you use the **auto** mode as the fan speed adjustment mode.

Table 19-59 lists the default configuration of the fan EMU.

Table 19-59 Default configuration of the fan EMU

Parameter	Default Value
Sub-node	1
Fan speed adjustment mode	Automatic
Whether to report the fan alarm	Permit

 Copyright © Huawei Technologies Co., Ltd.

Procedure

Step 1 Set the DIP switches of the sub-nodes for the fan EMU. By default, the sub-node ID is 1.

NOTE

The fan EMU communicates with the MA5800 in the master node and sub-node mode. Therefore, the DIP switches of the sub-nodes for the fan EMU must be consistent with those for the MA5800.

Step 2 Insert the fan tray into the corresponding slot of the service shelf.

Step 3 Run the **emu add** command to add a fan EMU. By default, the sub-node ID is 1.

Step 4 Run the **interface emu** command to enter the FAN mode.

Step 5 Run the **fan speed mode auto** or **fan speed mode manual** command to set the fan speed adjustment mode. By default, the fan speed adjustment mode is automatic.

NOTE

When the fan speed adjustment mode is the manual mode, you can run the **fan speed mode manual** command to set the fan speed. The speed level can be 0, 1, 2, 3, 4, 5 or 6. Here, 6 stands for the highest speed, and 0 stands for the lowest speed.

Step 6 Run the **fan alarmset** command to configure the fan alarm reporting function. The fan alarms are read temperature failure alarm, fan block alarm, over temperature alarm, and power fault alarm. By default, the fan alarm reporting is permitted.

Step 7 Run the **save** command to save the data.

----End

Result

- In the FAN mode, run the **display fan system parameter** command to query the parameter configuration of the fan tray and ensure that the configuration is the same as the data plan.
- In the FAN mode, run the **display fan environment info** command to query the running status of the fan tray and ensure that it is the same as the data plan.
- In the FAN mode, run the **display fan alarm** command to query the alarm information reported by the fan tray. The status of all the fan alarms is normal.

Example

To add a fan EMU, and adopt the default settings for the speed adjustment mode and alarm function, do as follows:

```
huawei(config)#emu add 0 FAN 0 1 FAN

huawei(config)#interface emu 0

huawei(config-if-fan-0)#display fan system parameter
```

 Copyright © Huawei Technologies Co., Ltd.

```
EMU ID: 0
FAN configration parameter:
--------------------------------------------------------------------------------
FAN timing mode: Auto timing by temperature
--------------------------------------------------------------------------------
Alarm_name                      Permit/Forbid
Read temperature fault                Permit
Fan block                       Permit
Temperature high                     Permit
Power fault                      Permit
--------------------------------------------------------------------------------
```

19.13 RTU

Definition

Right to Use (RTU) is a part of hardware and is used to authorize the use of hardware capabilities in phases, such as ports, bandwidth and optical power. RTU is a license for the use of hardware resources and has the same lifecycle as hardware.

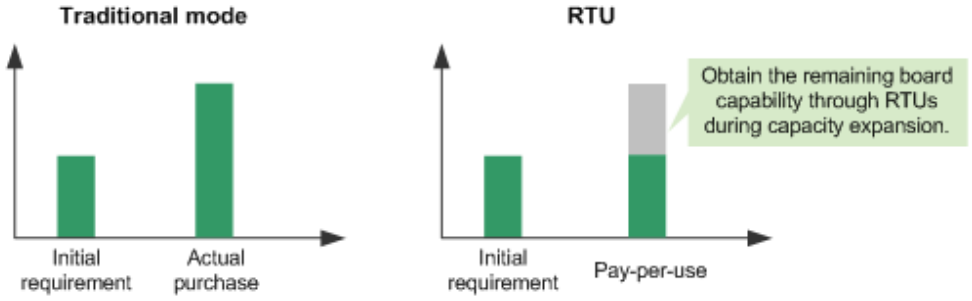
Carrier

To provision services using RTUs, the following requirements must be met:

1. Hardware: service subracks, boards , and optical modules that support RTUs.
2. Software: software version that supports the RTU, including the host version and NMS version.
3. RTU authorization certificate: type and number of RTUs that have been applied for.

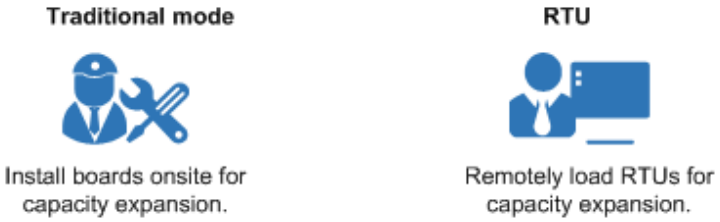
Benefits

- Reduced initial investment: The RTU provides customers with a method of purchasing hardware capabilities in installments, which can significantly reduce initial investment and quickly achieve balance between income and expenditure.



- Reduced OPEX: Services can be provisioned by remotely loading RTUs. Therefore, customers do not need to install boards onsite to achieve capacity expansion.

 Copyright © Huawei Technologies Co., Ltd.



RTU Type

Table 19-60 RTU types supported by access products

RTU Type	Function
Port RTU	Controls the number of available ports based on port access modes, including GPON, 10G PON, and so on.
Optical power RTU	Controls the number of ports whose working mode is enhanced based on the optical module type, such as GPON Class C+ Pro and 10G PON Class C+ Pro optical modules.
Bandwidth capability RTU	Controls the number of ports with specific access bandwidth capabilities, such as 10G PON initial bandwidth, 2.5G upstream bandwidth, and 2.5G downstream bandwidth capability, based on XGS PON/XGS combo PON physical ports.

For more information, see the following:

- Carrier users: **Access Network Product RTU**.
- Enterprise users: **Access Network Product RTU**.

 Copyright © Huawei Technologies Co., Ltd.

20 Slicing

Slicing is to divide a physical network into multiple logical networks for multiple purposes. Different slices can meet differentiated SLA, isolation, security, and independent operation requirements, effectively solving network congestion problems.

20.1 Slicing Solution Change Description
This section describes the changes in each version.

20.2 Slicing Solution Overview
This section describes the background, principle architecture, and typical application scenarios of the slicing solution.

20.3 Slice Lifecycle Management
On an actual network, slices are dynamically created and changed based on user requirements.

20.4 Slice Deployment and Configuration
This section describes the common slice deployment and configuration procedures and examples.

20.5 Reference

20.1 Slicing Solution Change Description

This section describes the changes in each version.

Release	Description
V100R023C11&C12	Optimized **20.4.1.2 Example of Hard-Isolated Dedicated Network Slice Deployment (CLI)** and added the procedure for adding an upstream port to a VLAN.
V100R023C11&C12	Added key technology **HDBA**.

 Copyright © Huawei Technologies Co., Ltd.

Release	Description
V100R023C00	• Optimized **20.4.1.2 Example of Hard-Isolated Dedicated Network Slice Deployment (CLI)** and added examples of configuring dedicate network slices by ONT and ONT port. • Optimized section *Prerequisites* and some section titles.
V100R021C10SPC300	Added the function of configuring MAC addresses for dedicated network slices. Added the link to P2MP dedicated network slicing.
V100R021C10	Added the function of dedicated network slicing.
V100R021C00	First release that supports service slicing.

For details about specifications and changes in each version, see **Specifications Query Tool**.

20.2 Slicing Solution Overview

This section describes the background, principle architecture, and typical application scenarios of the slicing solution.

20.2.1 Background of Slicing

With the improvement of living standards, users' network requirements become more and more diversified, and the requirement for fast construction of dedicated networks becomes a trend.

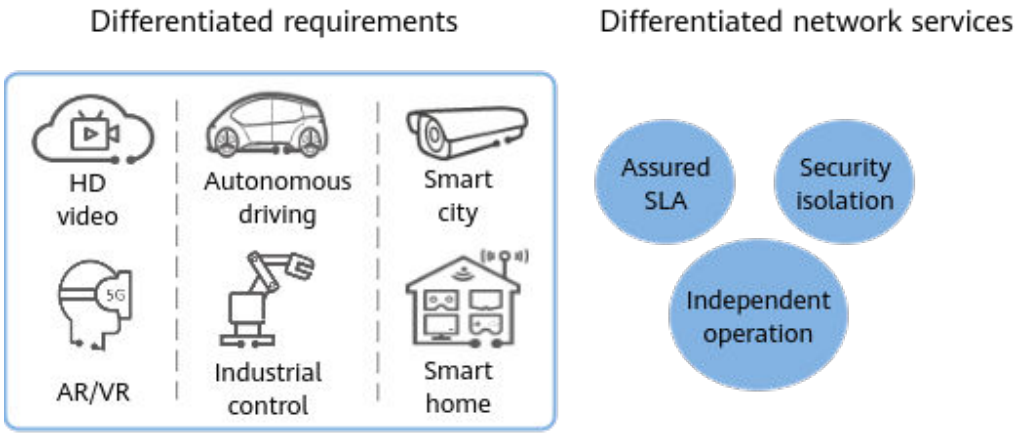
Diversified User Services and Differentiated User Experience

With the development of science and technology and the improvement in living standards, home and enterprise users have higher requirements on network quality. They are no longer satisfied with mere improvement in broadband rates, and are calling for more diversified experience. Communication is changing from connections between people to connections between people, between people and things, and between things. Internet of everything (IoE) has become a trend.

In the era of IoE, massive devices need to access the network, and each device may carry multiple services. These devices and services belong to different industries and have different characteristics, posing different requirements on network security, latency, reliability, and bandwidth. For example:

- Entertainment services such as HD video and AR/VR require high-bandwidth network connections.
- Autonomous driving and industrial control require low-latency and high-reliability network connections.
- Smart cities and smart homes require a large number of device connections and frequent transmission of small data packets.

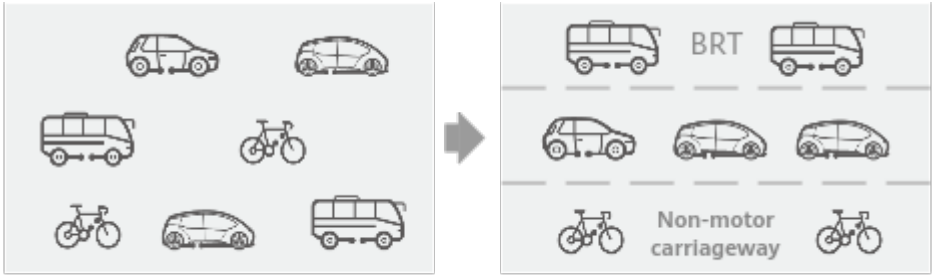
 Copyright © Huawei Technologies Co., Ltd.



Differentiated Network Service Assurance by Operators

A communications network is like a transportation system. Data packets are "vehicles", and the network provides "roads". Before cars became popular, the main consideration was to make roads wider and flatter to meet the needs of motorcycles and bicycles. As the number of vehicle types increases, the requirements for roads are also changing. It is not reasonable to use the same road and network mechanism to support bicycles and sports cars. To solve the problem of traffic jams during rush hours in the morning, it is not economical to only expand roads. The transportation department also controls traffic based on vehicle priorities and operation modes. For example, the bus rapid transit (BRT) lane and non-motor lane are set to improve the traffic efficiency.

Figure 20-1 Road traffic distribution management



Similarly, when network services are relatively simple and bandwidth is the main indicator, operators only need to consider how to improve the network speed and stability. In the era of IoE, numerous devices and applications are connected to the network. Home users and industry users have different and customized requirements. One network or the same set of policies cannot meet the requirements of new services and scenarios. Independent network construction for each service scenario is costly, time-consuming, and uneconomical.

- In multi-service unified bearing scenarios, for example, in home broadband, one home terminal may access multiple applications (such as online courses, online games, live broadcast, online shopping, and online chat) at the same time. These services have different network assurance requirements and priorities. Operators need to provide differentiated assurance for different services.

 Copyright © Huawei Technologies Co., Ltd.

- In multi-tenant independent operation scenarios, if an independent hardware infrastructure network is provided for each network service provider (NSP)/ Internet service provider (ISP), the cost is high and device usage is low. If physical network resources can be shared by multiple tenants, the investment will be significantly reduced, and the device usage will be higher.

Slicing Solution

Based on the current software defined networking (SDN) technology, Huawei has developed a technology to divide one hardware infrastructure network into multiple independent virtual networks, for which resources can be flexibly combined and assigned on demand to meet different service requirements. If the live network cannot meet a new service requirement, an operator can create a new virtual network without affecting the live network. In this way, the new service can be provisioned as soon as possible.

The slicing solution helps operators and ISPs build networks more efficiently and agilely and provide differentiated and isolated network services for users.

- It helps operators provide differentiated service experience for home broadband users and provide isolated and secure logical dedicated networks for industry users.
- It also enables ISPs to obtain isolated access network resources economically and carry out independent service deployment and O&M.

20.2.2 Benefits of the Slicing Solution

Network slicing provides resource isolation, differentiated service level agreements (SLA), high reliability, flexible topology customization, and automatic slice management capabilities, helping enterprises achieve digital transformation.

Network slicing enables operators to build multiple dedicated, virtual, isolated, and on-demand logical networks on one physical network to meet network capability requirements (such as latency, bandwidth, and number of connections) from different industry customers.

- **Differentiated SLAs for diversified services**. Customized slices based on service requirements enable networks to have good service adaptability and meet differentiated service requirements such as latency, jitter, reliability, and security. Therefore, the slicing technology can help operators enter the massive vertical industry markets.
- **Dynamic resource assignment and flexible networking**. Through network slice life cycle management, the service, function, capacity, service quality, and connection relationship of a network slice can be customized based on different service requirements, implementing dynamic assignment and efficient use of network resources.
- **One network for multiple purposes, reducing costs.** Compared with independent networks, network slicing enables a unified infrastructure network to meet multi-service requirements, avoids the need to build a dedicated physical network for each service, reduces network construction investment, improves the network operation efficiency, and reduces operators' CAPEX and OPEX.

 Copyright © Huawei Technologies Co., Ltd.

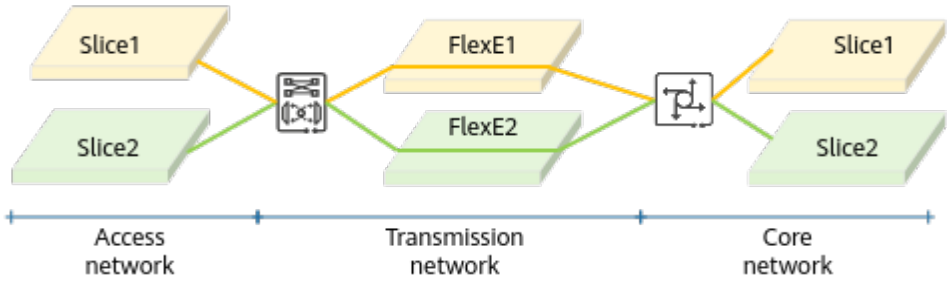
- **Simplified configuration.** ONU services can be automatically configured in batches without the need for onsite software debugging, improving device deployment and service provisioning efficiency.
- **Visualized O&M.** The network management system (NMS) visualizes slice resources and slice SLAs, and monitors network-wide status in real time, so that users can view the real-time running status of the entire network clearly.

20.2.3 Architecture of the Slicing Solution

The network slicing solution is an end-to-end solution that flexibly assigns resources and combines network capabilities based on service requirements from the access network, transmission network, to the core network. The network slice management system is used to implement unified network management.

E2E Slicing

To implement end-to-end network slicing, each network segment must support network slicing and identify network slices in an end-to-end manner, as shown in the following figure.



The basic slice information of each NE is as follows:

Access network

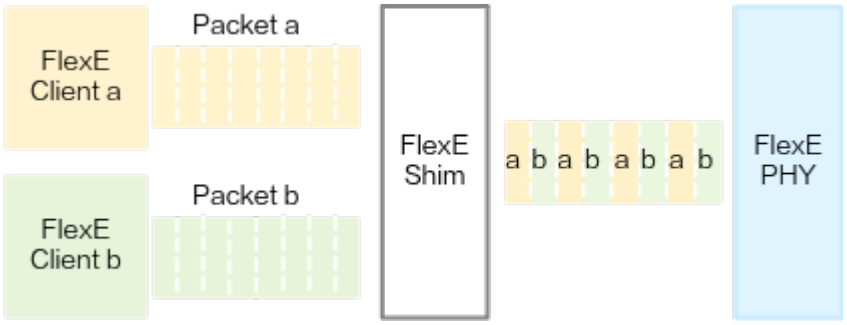
Slicing object: port, ONU, and transmission bandwidth

Slicing mode: service-based soft slicing and port-based hard slicing

Transmission network

Slicing object: transmission bandwidth

Slicing mode: differentiated hard pipe isolation using FlexE



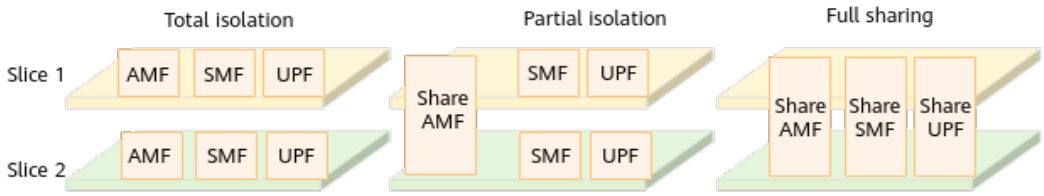
Core network

 Copyright © Huawei Technologies Co., Ltd.

Slicing object: virtualized network function (VNF)

Slicing mode: VNF sharing planned based on the SLA requirements of services. Different VNF sharing policies lead to different network isolation, service performance, and networking complexity.

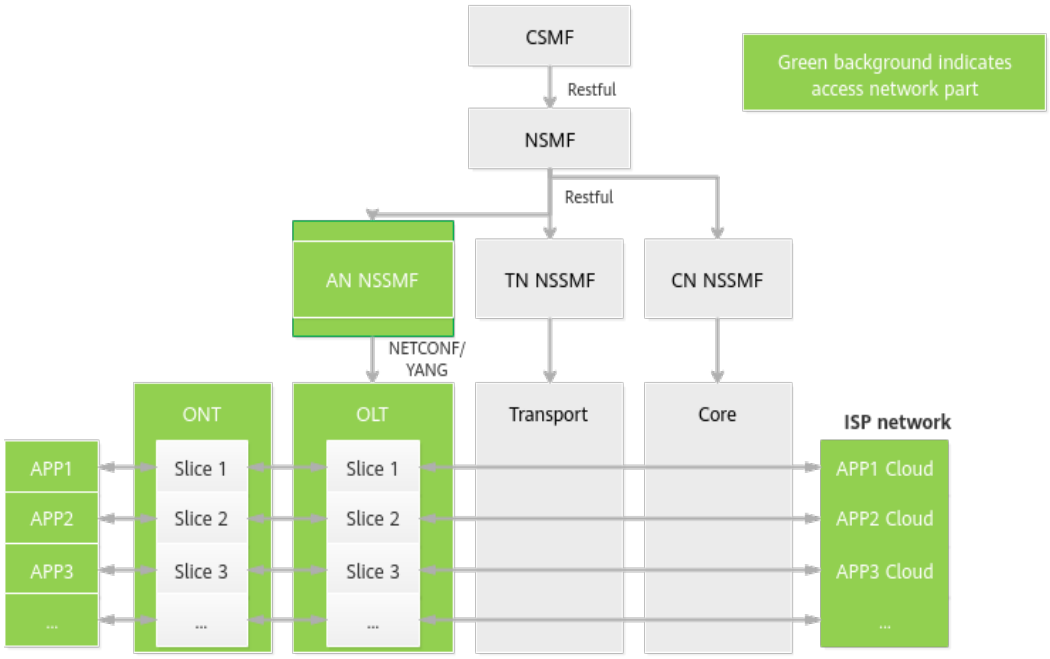
- VNF that can be shared by all network slices: network slice selection function (NSSF), network repository function (NRF), and network exposure function (NEF).
- VNF that can be shared by some network slices: access and mobility management function (AMF), unified data management (UDM), and policy control function (PCF).
- VNF that can be exclusively used by a network slice: session management function (SMF) and user plane function (UPF).



Architecture of the Slicing Solution

Figure 20-2 shows the architecture of the network slicing solution and protocols used by main interfaces.

Figure 20-2 Architecture of the network slicing solution



The communication service management function (**CSMF**) converts service requirements into network requirements. The CSMF provides a slice mall. After a user subscribes to a slice service in the mall, the CSMF converts the service

 Copyright © Huawei Technologies Co., Ltd.

requirements of the user into requirements on the network side and transfers the requirements to the NSMF for network design. Generally, the CSMF is provided by the business support system (BSS) of an operator.

The network slice management function (**NSMF**) manages and orchestrates end-to-end network slices. After receiving end-to-end network slice requirements, the NSMF converts the requirements into slice instances, decomposes and combines services based on the capabilities of each subdomain/subnet (including the access network, transmission network, and core network), and transfers the deployment requirements on subdomains/subnets to the NSSMF.

The NSMF is generally provided by a cross-domain slice manager.

The network slice subnet management function (NSSMF) manages and orchestrates network slice subnet instances (NSSIs), including slice life cycle management (such as deployment, activation, and deactivation) and O&M (such as slice subnet performance monitoring and alarms). It focuses on a single subnet, such as the access network.

Communication Interfaces

As shown in **Figure 20-2**, the communications interfaces between components in the slicing solution are as follows:

CSMF and NSMF: RESTful protocol

NSMF and NSSMF: RESTful protocol

NSSMF and access NEs: NETCONF/YANG

NOTE

This document describes the implementation, application, and configuration of the access network in detail. For details about the transmission network and core network, see the corresponding product documentation.

20.2.4 Implementation Principles of Access Network Slicing

Access network slices can be classified into dedicated network slices and service slices.

dedicated network slice: Divides software and hardware resources of a network into multiple independent slices. Each slice forms a hard isolated dedicated network. Different virtual network operators (VNOs) perform independent and isolated O&M management on dedicated network resources. It also enables a campus network to carry multiple subnet services in a secure manner.

Service slice: Divides a traditional access network or dedicated network into service channels of different SLAs. Different network slice instances (NSIs) are deployed in service channels to meet differentiated network requirements of different users or services.

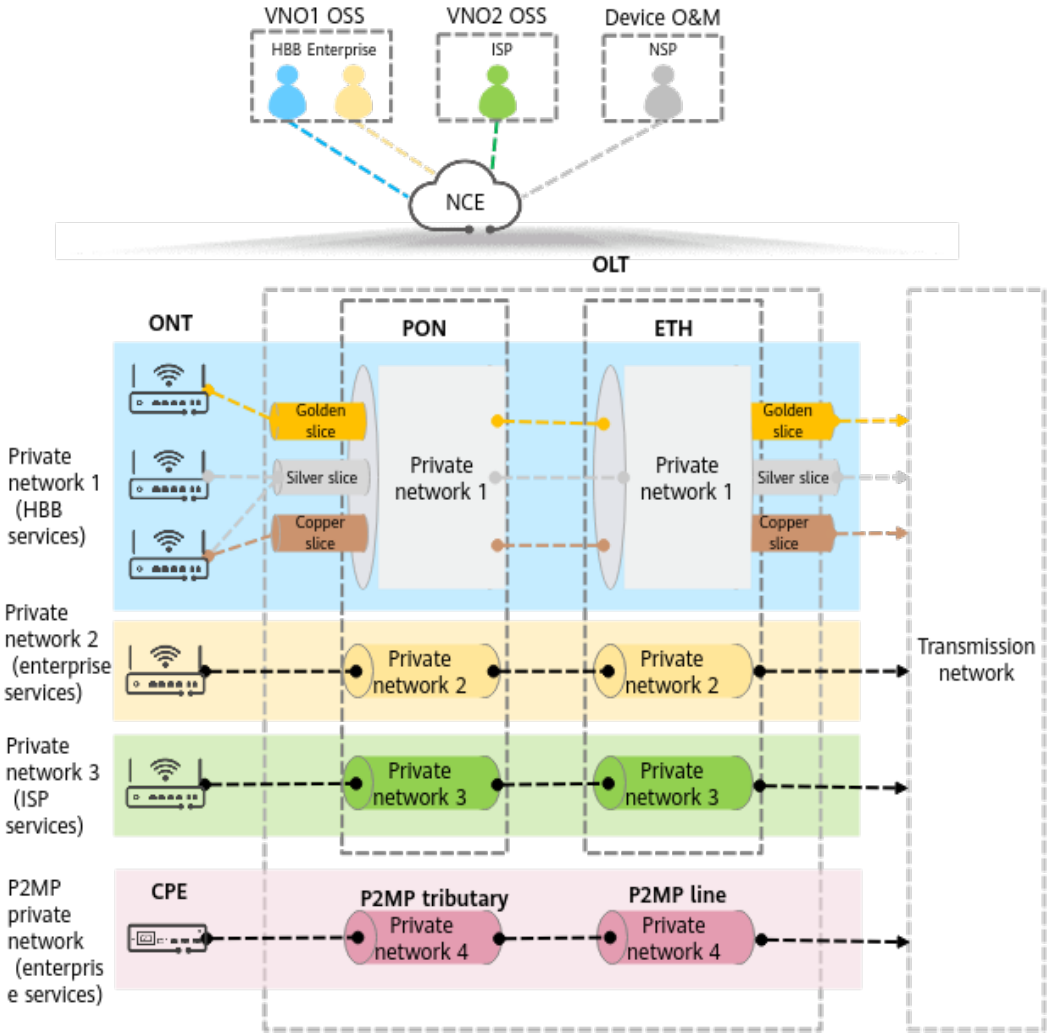
 Copyright © Huawei Technologies Co., Ltd.

Table 20-1 Comparison between the two slicing modes

Item	Dedicated Network Slice	Service Slice
Object	Deterministic services	General services
Slicing granularity	Networking level	Service level
Slice level classification	Not classified	3 classes: golden, silver, and copper
Application scenario	Enterprises requiring private lines or ISP leasing, enterprise campus networks, and power distribution networks	Differentiated SLAs for multiple services on one network (for example, multiple services in home broadband or enterprise scenarios)
Service deployment modes	On-demand deployment	Pre-deployment
O&M and management	Isolated	Unified NMS and independent slice view

 Copyright © Huawei Technologies Co., Ltd.

Figure 20-3 Access network slicing principle



Dedicated Network Slice

An NSP uses the NCE controller to slice OLT resources into multiple dedicated networks. Resources and services of the dedicated networks are isolated from each other. The NSP can lease network slices to ISPs. Each ISP performs isolated O&M through an independent VNO interface. The logical topology, route selection, and O&M of each VNO are independent of those of other VNOs.

- **NSP multi-tenant management**: The NSP uses the NMS to manage tenants in a visualized and unified manner and assign network resources.
- **Resource and service isolation between ISP tenants**: By slicing resources at the NCE layer, an NSP can assign the port, VLAN, and other resources of an OLT to ISPs so that the ISPs can share the OLT and fiber resources. The rights control on the tenant interface enables independent management of resources between ISPs and improves the reliability of private line services.
- **Independent service provisioning and O&M by ISPs**: NCE provides different service package profiles for tenants to implement differentiated services between ISPs and meet different service requirements. The one-click troubleshooting function provided by NCE helps ISPs quickly diagnose faults.

 Copyright © Huawei Technologies Co., Ltd.

dedicated network slicing includes two solutions.

- Physically dividing the PON board resources of an OLT
- Building E2E P2MP dedicated networks using P2MP boards

Service Slices

Network service experience requirements are classified into the following types:

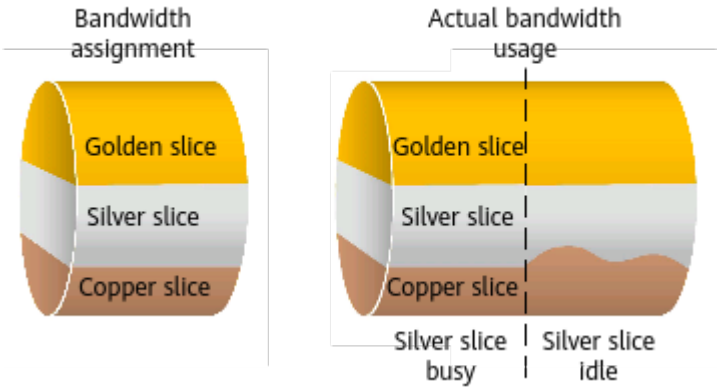
- Golden services: services that require assured bandwidth and latency
- Silver services: services that require assured bandwidth
- Copper service: regular services

Based on network service experience requirements, access network service slices are classified into golden slices, silver slices, and copper slices. The characteristics of resource assignment are as follows:

- Golden slice: fixed reserved bandwidth which cannot be preempted even when idle
- Silver slice: assured bandwidth which can be preempted when idle
- Copper slice: best-effort bandwidth, allowing overselling

In each type of golden, silver, and copper slices, multiple specific service slices with different network parameters can be set. For example, in the golden slice type, a VR service slice, an HD video service slice, and a live broadcast service slice can be set.

Figure 20-4 Bandwidth usage supported by golden, silver, and copper slices



 Copyright © Huawei Technologies Co., Ltd.

Table 20-2 Comparison of golden, silver, and copper slice resources

Slice Category	Direction	ONT	
		Wi-Fi Side	WAN Side
Golden slice	Upstream	• Exclusive RU resources • Dynamic assignment	• Application identification • Virtualization into multiple WAN ports that correspond to different slices
	Downstream		
Silver slice	Upstream	• RU resource sharing • Scheduling with priority higher than copper slices	
	Downstream		
Copper slice	Upstream	RU resource sharing	
	Downstream		

Slice Category	Direction	OLT	
		PON	Uplink
Golden slice	Upstream	• DBA fix: fixed bandwidth that cannot be preempted even when idle • T-CONT timeslot isolation	• Highest-priority HQoS scheduling • Fixed bandwidth that cannot be preempted even when idle • Independent slice queues and configurable packet buffer size
	Downstream	• Highest-priority HQoS scheduling, and fixed bandwidth that cannot be preempted even when idle • Independent slice queues and configurable packet buffer size	/

 Copyright © Huawei Technologies Co., Ltd.

Slice Category	Direction	OLT	
		PON	Uplink
Silver slice	Upstream	• DBA assure: assured bandwidth that can be preempted when idle • T-CONT timeslot isolation	• Medium-priority HQoS scheduling • Assured bandwidth that can be preempted when idle • Independent slice queues and configurable packet buffer size
	Downstream	• Medium-priority HQoS scheduling, and assured bandwidth that can be preempted when idle • Independent slice queues and configurable packet buffer size	/
Copper slice	Upstream	• DBA Max: maximum bandwidth • T-CONT timeslot isolation	• Low-priority HQoS scheduling • Bandwidth overselling allowed, best effort

Slice Category	Direction	MAN
Golden slice	Upstream	OTN+ preferred VXLAN and VLAN supported
	Downstream	
Silver slice	Upstream	VXLAN preferred OTN+ and VLAN supported
	Downstream	
Copper slice	Upstream	VLAN preferred OTN+ and VXLAN supported

20.2.5 Key Technologies

20.2.5.1 HDBA

DBA Principles

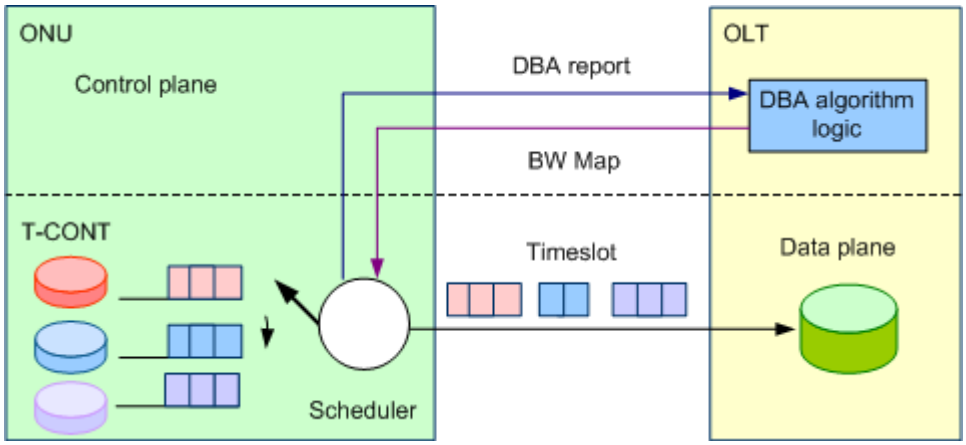
DBA is short for dynamic bandwidth assignment. An OLT uses DBA to dynamically adjust the upstream bandwidth assigned to different ONUs to process the burst

 Copyright © Huawei Technologies Co., Ltd.

traffic from ONUs, meeting the ONU upstream bandwidth requirements and improving the utilization of the PON upstream bandwidth.

Figure 20-5 shows the DBA principles.

Figure 20-5 DBA principles



DBA implementation process:

- The DBA module in an OLT consistently collects DBA reports and uses a DBA algorithm to calculate the upstream bandwidth assigned to each ONU.
- The OLT sends the calculated result to each ONU using a bandwidth (BW) map.
- Each ONU transmits upstream data in burst mode in permitted timeslots defined in the BW map.

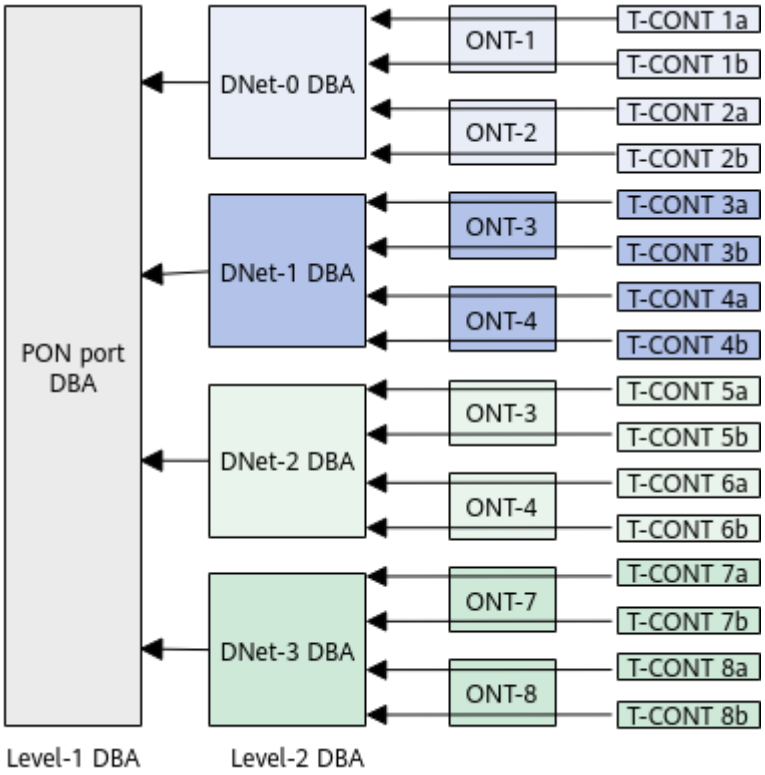
HDBA Principles

When multiple operators share an ODN, users connected to the same PON port can select service packages provided by different operators. A unique dedicated network (DNet) can be assigned to each operator, so that each operator can provision services and perform O&M separately. When dynamically assigning bandwidth, an OLT uses the hierarchical DBA (HDBA) technology to add the DNet group attribute to a PON port and performs 2-level DBA scheduling to implement bandwidth sharing on the PON port.

HDBA refers to two-level DBA scheduling by adding a DNet group attribute to a PON port. During dynamic bandwidth assignment, the bandwidth scheduling on a PON port is changed from T-CONT-based scheduling to two-level scheduling from DNet groups to T-CONTs. **Figure 20-6** shows the HDBA principles.

 Copyright © Huawei Technologies Co., Ltd.

Figure 20-6 HDBA Principles



20.2.6 Typical Application Scenarios of the Slicing Solution

Network slicing enables traditional physical network resources to be flexibly and precisely divided, and slices are isolated from each other. It provides differentiated service experience for home and enterprise users, and provides ISPs with more flexible and economic physical networks.

20.2.6.1 Application Networking of the Slicing Solution

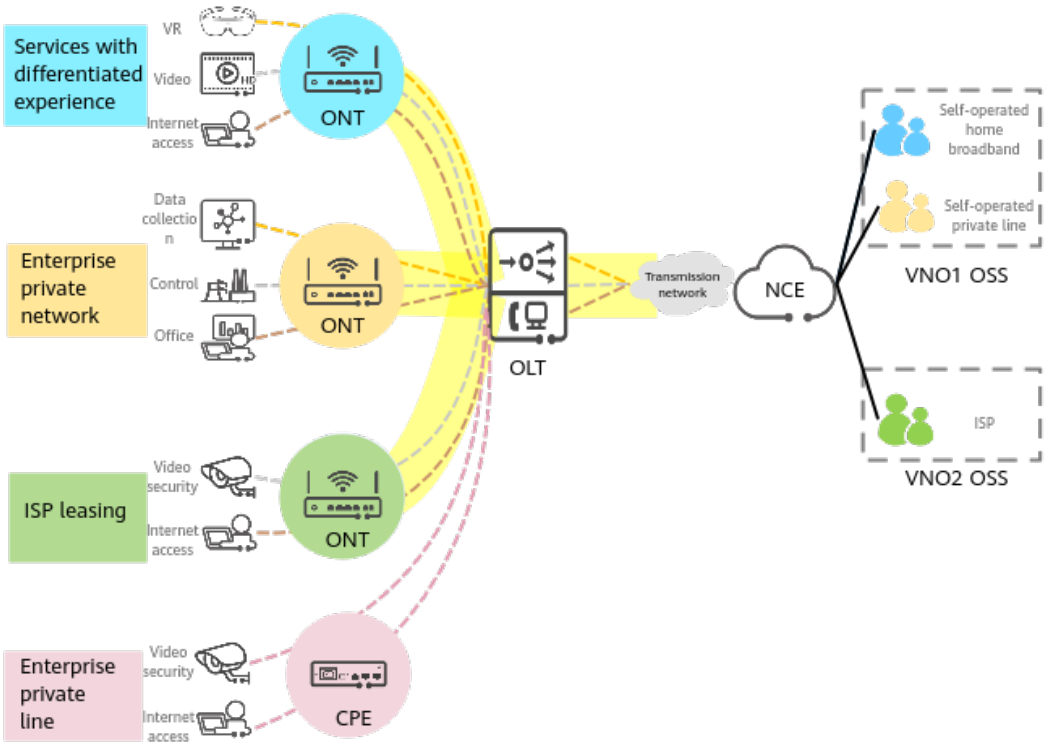
The access network slicing solution can provide differentiated services on the same FTTH network. It can also be used to build dedicated networks for enterprises with high requirements or lease the network resources to ISP users.

Figure 20-7 shows the usage of the two slicing modes on the access network.

- Service slice: applies to home broadband scenarios or enterprise scenarios where different services require differentiated network services. For example, differentiated management can be implemented for VR, video, and common Internet access services in home broadband, or for data collection, industrial control, and office services in enterprises.
- dedicated network slice: Applies to independent O&M of multiple VNOs, such as enterprise private lines or ISP dedicated networks, and the scenario of one campus network carrying multiple subnet services in a secure manner.

 Copyright © Huawei Technologies Co., Ltd.

Figure 20-7 Slice application networking



20.2.6.2 Customized Network Services for Dedicated Network Slices

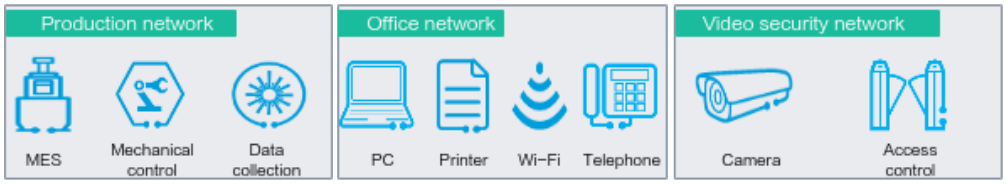
dedicated network slices apply to scenarios such as independent O&M by multiple VNOs and enterprise private lines.

Generally, for customers such as government departments, enterprises, and healthcare, finance, energy, and electricity industry customers, there are multiple service networks on their local area networks (LANs), including office and production networks on the intranet, Internet access, Wi-Fi, Internet of Things (IoT), and device networks (security, video, and access control). These services are isolated through dedicated network slicing to meet user requirements for high security.

Campus Scenarios

Enterprise campuses can use dedicated network slicing to isolate different services.

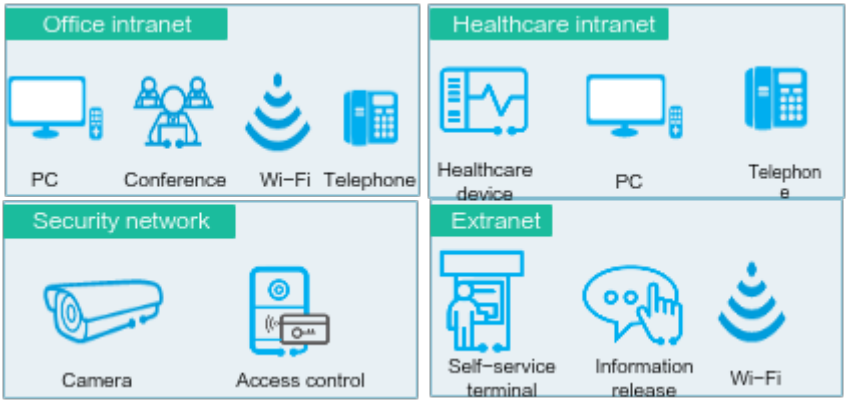
- Industrial campuses can use dedicated network slicing to isolate their production, office, and security networks.



- On a hospital campus, there are usually three networks — the intranet, device network, and extranet. Some hospitals further divide their intranets into office

 Copyright © Huawei Technologies Co., Ltd.

and healthcare intranets. Hospital campuses can use dedicated network slicing to isolate the preceding networks.

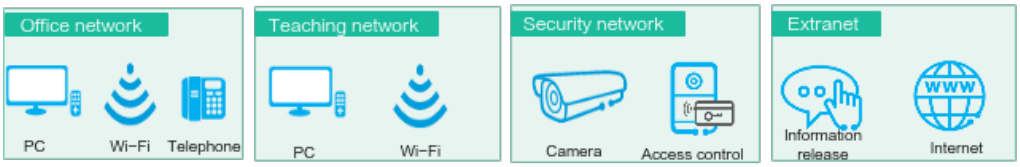


The **hospital intranet** is a dedicated medical information network that is open only to internal staff. It is the core working network of a hospital. The hospital intranet carries smart healthcare data and most smart management services, such as the hospital information system (HIS), electronic medical record (EMR), laboratory information system (LIS), picture archiving and communication system (PACS), and hospital resource planning (HRP) system. As the core network, the hospital intranet is also used for Wi-Fi access and medical IoT, and covers the consultation areas and wards of each clinical department, medical technology department, office area, cashiers, pharmacies, outpatient, emergency halls, and public areas.

The **hospital device network** is used to connect security cameras, environment monitoring devices, and access control systems. As safety incidents occur due to medical disputes from time to time, a reliable and economical security system is increasingly significant for hospitals, especially in key areas of smart hospitals.

The **hospital extranet** is used for Internet and external Wi-Fi access in scenarios such as offices and public areas. It also provides services such as appointment and treatment, online follow-up consultation, information release, result check, and telemedicine. With the development of "Internet + healthcare", hospitals are providing smart services on the Internet as well, such as online reservation, consultation, and payment.

- School campuses can use dedicated network slicing to isolate their office, teaching, and security networks, and extranet.



Electricity Scenarios

On electricity transmission/distribution networks, there are control services and data collection service that need to be isolated from each other.

 Copyright © Huawei Technologies Co., Ltd.

ISP Scenarios

Dedicated network slices can be used when multiple ISPs share resources of one physical network.

Table 20-3 Network requirements of different services

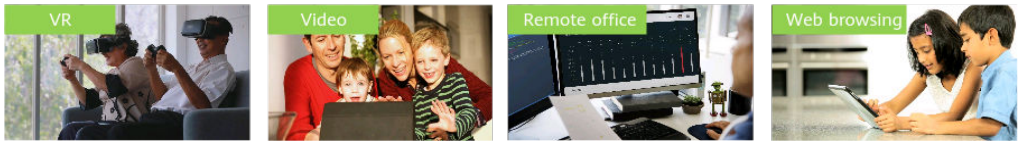
Service Type	Safe City	Finance	Energy	ISP Self-Operation
Network requirements	High isolation High reliability High bandwidth Regular latency Massive connections	High isolation High reliability High bandwidth Lower latency	High isolation High reliability Lower latency	Customized requirements

20.2.6.3 Differentiated Service Experience of Service Slices

Service slices apply to home broadband scenarios or enterprise scenarios where different services have differentiated network requirements.

Home Broadband Scenario

Services such as VR, video call, office, and web browsing in home broadband have different network requirements, and differentiated SLA management can be provided.



 Copyright © Huawei Technologies Co., Ltd.

Table 20-4 Network requirements and slice recommendation for different services

Service Type	VR	Video and Remote Office	Web Browsing
Network requirements	Strong interactivity Bandwidth: high Latency: low Jitter: small Packet loss: low	Real-time Bandwidth: average Latency: relatively low Jitter: relatively small Packet loss: relatively low	Regular network requirements
Recommended slice	Golden slice	Silver slice	Copper slice

Enterprise Scenario

An enterprise campus network is divided into different areas or data groups with different network requirements. For example, production data collection, industrial control, online training, and office require differentiated management.

Table 20-5 Network requirements and slice recommendation for different services

Service Type	Production Center	Training	R&D Office/ Administrative Office
Network requirements	Large data volume and high reliability requirements Bandwidth: high Latency: low Jitter: small Packet loss: few	Strong interactivity Bandwidth: average Latency: relatively low Jitter: relatively small Packet loss: relatively low	Regular network requirements
Recommended slice	Golden slice	Silver slice	Copper slice

 Copyright © Huawei Technologies Co., Ltd.

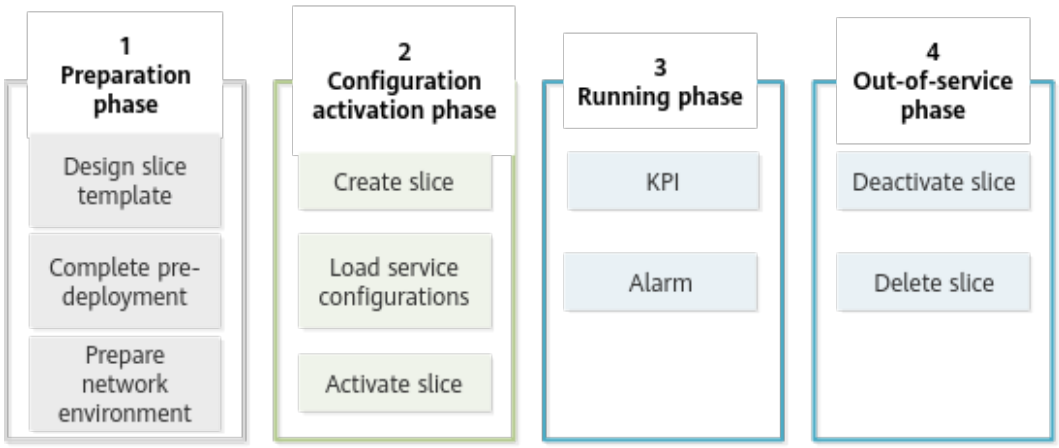
20.3 Slice Lifecycle Management

On an actual network, slices are dynamically created and changed based on user requirements.

When a user dynamically creates a network slice based on a network slice template, the network slice has a specific network capability of supporting a type of service, that is, the network slice becomes activated and alive.

Slice Life Cycle

The life cycle of a network slice is divided into four phases to provide more flexible services to tenants.

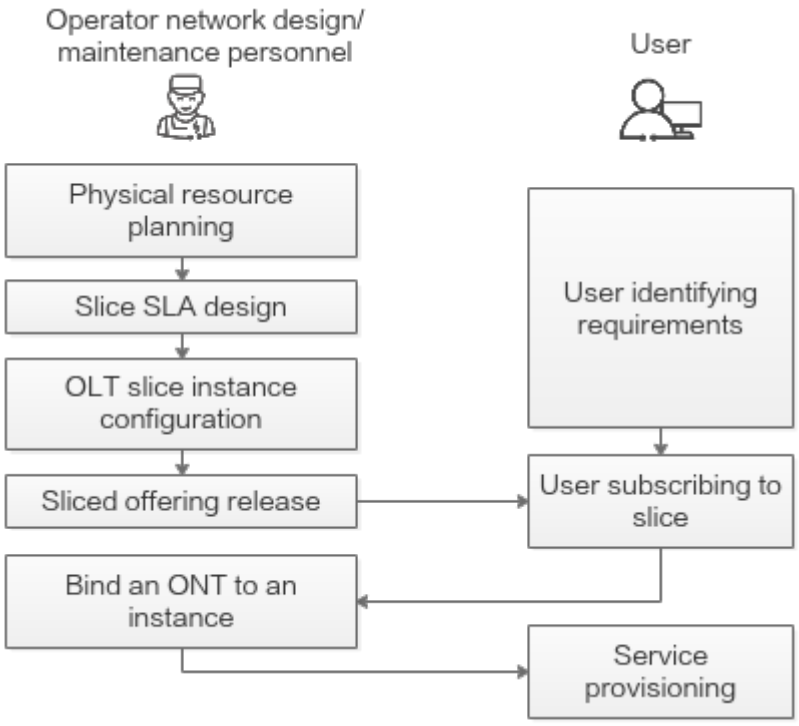


- **Preparation phase**: Design the slice template based on actual service requirements and complete pre-deployment to prepare the network environment.
- **Configuration activation phase**: Create slices, load service configurations, and activate slices based on users' subscription status to provide differentiated SLAs for users.
- **Running phase**: Monitor and manage resource usage and clear alarms during slice running.
- **Out-of-service phase**: Deactivate and delete a slice after a user unsubscribes from the slice or the slice plan is changed.

Slice Management Process

The following figure shows the overall slice provisioning process based on a service or application.

 Copyright © Huawei Technologies Co., Ltd.



20.4 Slice Deployment and Configuration

This section describes the common slice deployment and configuration procedures and examples.

20.4.1 Dedicated Network Slice Deployment

To implement independent operation and management of dedicated networks, tenants need to be configured, independent physical resources need to be configured for the tenants, and then the tenants can configure services on their own network management pages.

20.4.1.1 Example of Hard-Isolated Dedicated Network Slice Deployment (NCE)

This section describes how to deploy dedicated network slices on NCE-FAN.

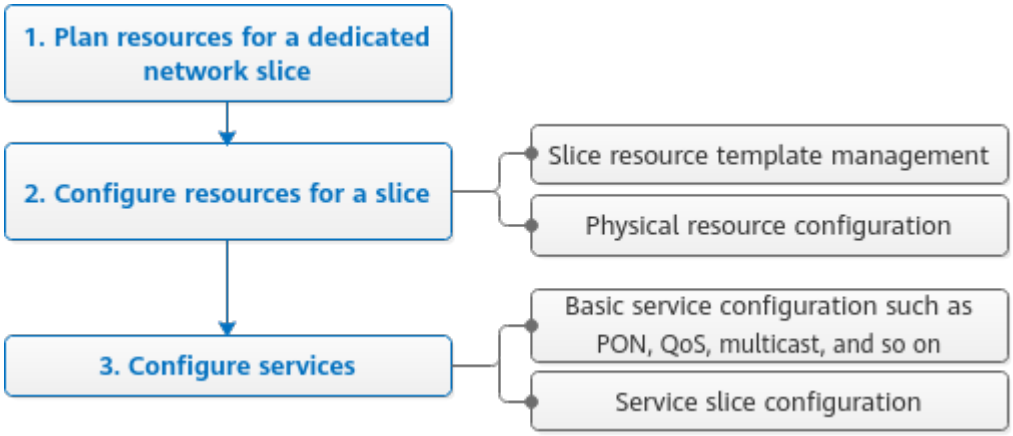
Prerequisites

- OLT model: MA5800-X17, X15, X7, and X2
- OLT version: V100R021C10 or later
- NCE-FAN version: V100R021C10 or later
- NEs have gone online, and the management channel for NEs has been established.

 Copyright © Huawei Technologies Co., Ltd.

- You have the permission to perform **network slice management**.

Dedicated Network Division and Configuration Process



Configuration Procedure

The GUI varies with the NMS version. For details about the operations in each version, see the configuration guide of the corresponding version.

Enterprise customers: **iMaster NCE-FAN Product Documentation**.

Carrier users: **iMaster NCE-FAN Product Documentation**

Data Planning Example in the iMaster NCE-FAN Product Documentation

Table 20-6 Key data planning

Parameter	Description	Example Value
Dedicated network ID	Indicates the ID of a dedicated network, which is automatically assigned by the OLT.	1
Dedicated network name	Indicates the name of a dedicated network	DNet-1
Upstream port	Indicates the upstream port bound to a dedicated network slice.	0/8/0
Service port	Indicates a service port bound to a dedicated network slice.	0/2/0
Service port bandwidth	Indicates the bandwidth of a service port, which is configured on a non-combo board.	300Mbit/s
Uplink port bandwidth	Indicates the bandwidth of the uplink port.	300Mbit/s

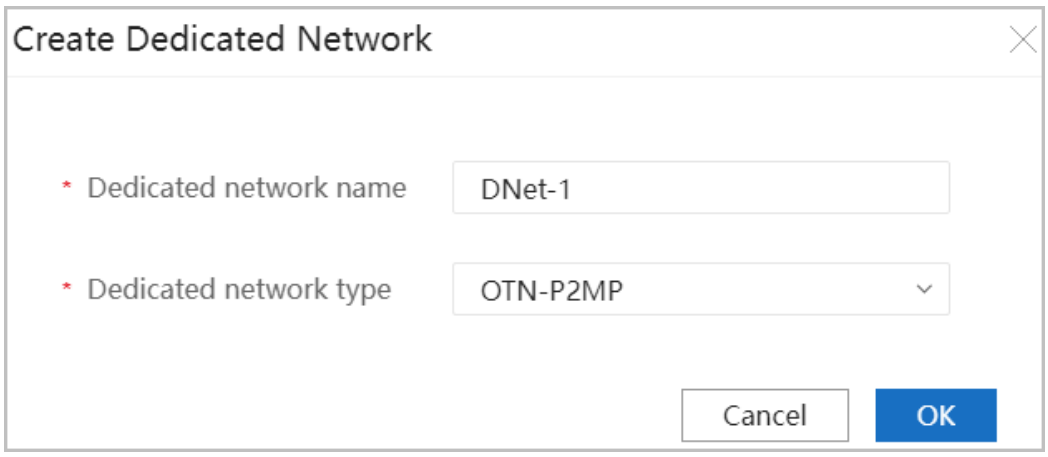
 Copyright © Huawei Technologies Co., Ltd.

Parameter	Description	Example Value
Channel 0 port bandwidth	Indicates the port bandwidth of channel 0. A combo board requires the channel parameter, whereas a non-combo board does not require the channel parameter.	300Mbit/s
Channel 1 port bandwidth	Indicates the port bandwidth of channel 1. A combo board requires the channel parameter, whereas a non-combo board does not require the channel parameter.	500Mbit/s

Procedure

Step 1 Open the **Access Controller** app and choose **Private Line** > **D-Net** from the main menu.

Step 2 Click **Create**. In the dialog box that is displayed, set **Dedicated network name** and **Dedicated network type**, and click **OK**.



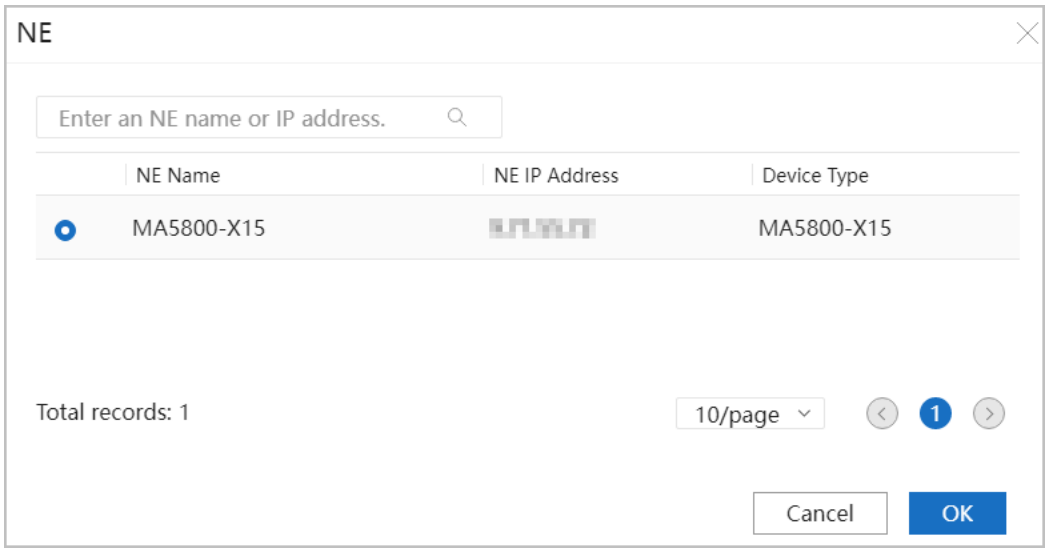
Step 3 Click > corresponding to the dedicated network slice to unfold the list of NEs bound to the slice.

Step 4 Click **Add**. In the dialog box that is displayed, select an NE and click **OK**.

NOTE

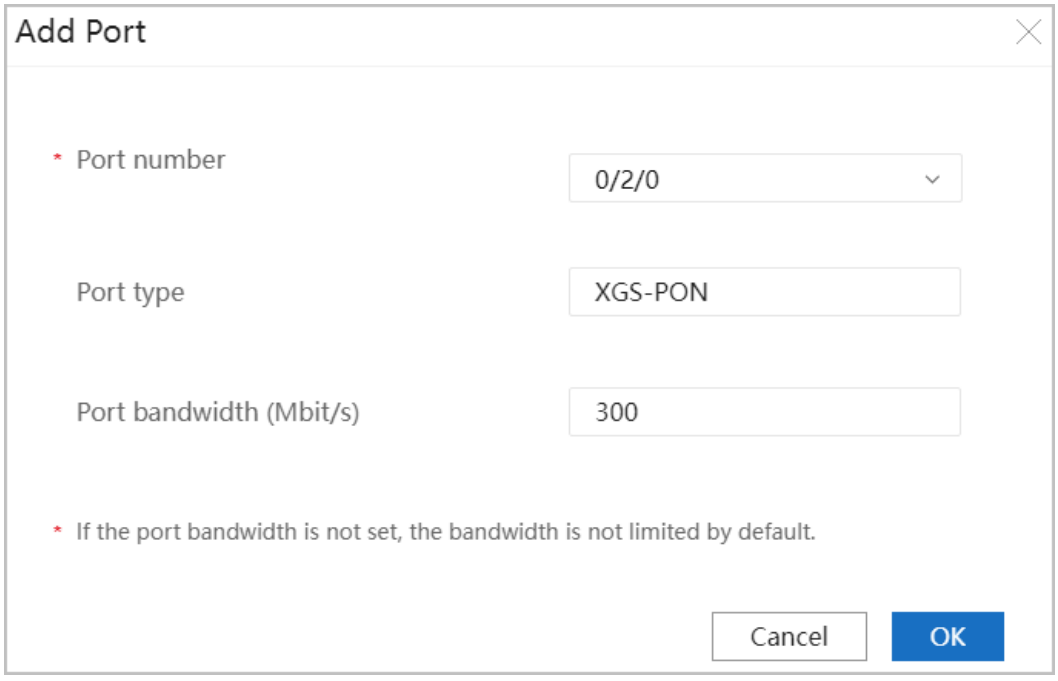
Each NE can be bound to a maximum of eight dedicated network slices, one of which is the default slice.

 Copyright © Huawei Technologies Co., Ltd.



Step 5 Click in the **Operation** column of an NE. In the dialog box that is displayed, configure port resources. The following uses the GPON port and upstream ETH port as examples.

- Configuring a PON port
  a. On the **GPON Port** tab page, click **Add**.
  b. In the dialog box that appears, set the parameters such as the port number and bandwidth.



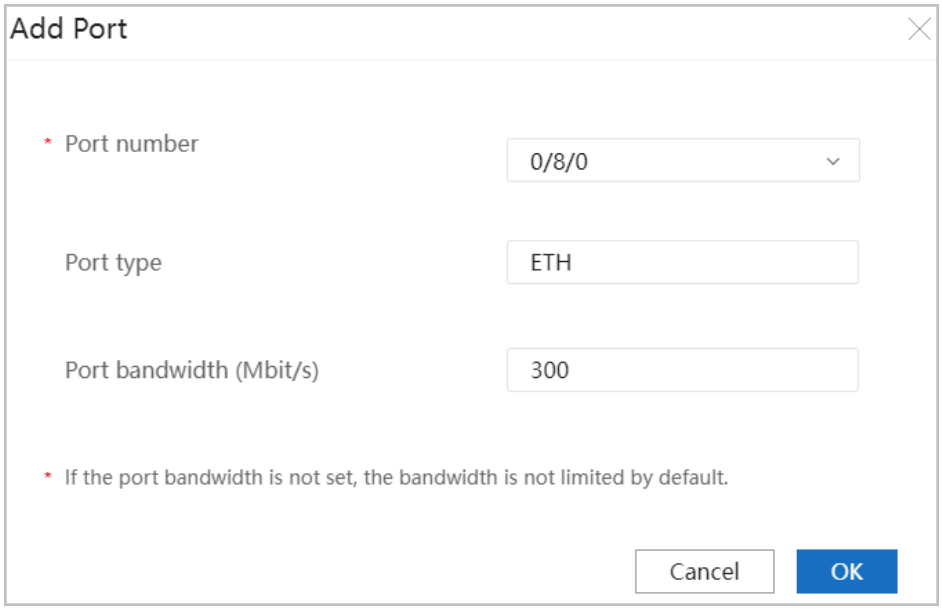
NOTE

If the PON port is a Combo port, you need to configure the channel port bandwidth.

  c. Click **OK**.

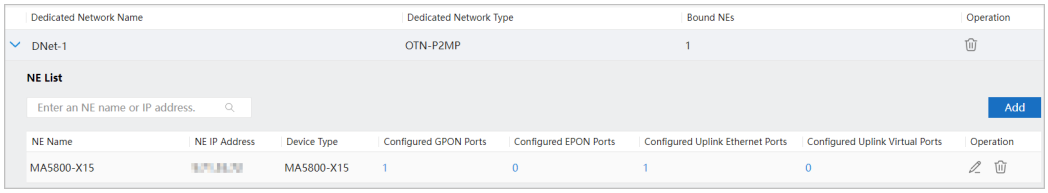
 Copyright © Huawei Technologies Co., Ltd.

- Configuring an upstream port.
  a. On the **Uplink Ethernet Port** tab page, click **Add**.
  b. In the dialog box that appears, set the parameters such as the port number and bandwidth.



  c. Click **OK**.

Step 6 After the configuration is complete, you can view the configured resource information.



----End

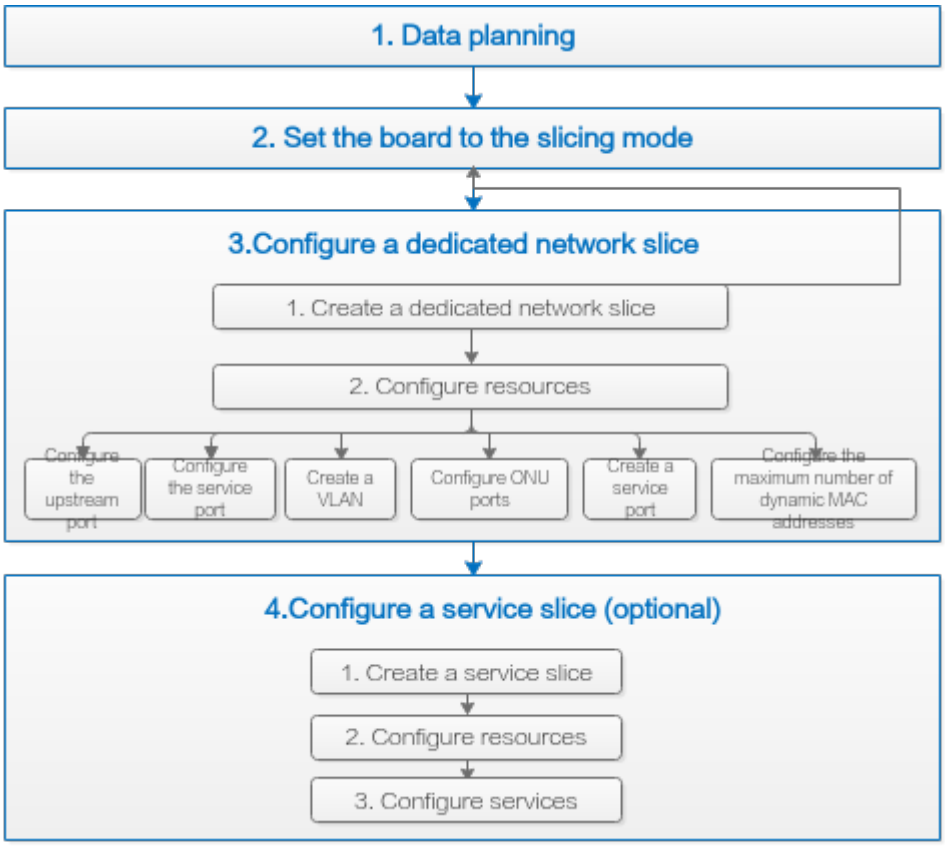
20.4.1.2 Example of Hard-Isolated Dedicated Network Slice Deployment (CLI)

Prerequisites

- OLT model: MA5800-X17, X15, X7, and X2
- OLT version: V100R021C10 or later
- NEs have gone online, and the management channel for NEs has been established.

 Copyright © Huawei Technologies Co., Ltd.

Configuration Procedure



Data Planning Example

Table 20-7 Data planning for dedicated network slices

Parameter	Description	Example Value
Dedicated network ID	/	1, 2
Dedicated network name	/	DNet-1 DNet-2
DNet-1 resource	OLT upstream port	0/9/0
	OLT service port	0/3/0
	PON port bandwidth channel 0 (A combo board needs to be configured with a channel, and a non-combo board does not need to be configured with a channel.)	700M

 Copyright © Huawei Technologies Co., Ltd.

Parameter	Description	Example Value
	PON port bandwidth channel 1 (A combo board needs to be configured with a channel, and a non-combo board does not need to be configured with a channel.)	7000M
	ONU resources	All ports of ONU 0 and ports 1–2 of ONU 1 connected to port 0/3/0
	VLAN	100
	DNet-1 GEM port index	1
	DNet-1 mapping index	0
DNet-2 resource	OLT upstream port	0/9/1
	OLT service port	0/3/0
	PON port bandwidth channel 0 (A combo board needs to be configured with a channel, and a non-combo board does not need to be configured with a channel.)	300M
	PON port bandwidth channel 1 (A combo board needs to be configured with a channel, and a non-combo board does not need to be configured with a channel.)	3000M
	ONU resources	Ports 3–4 of ONU 1 connected to port 0/3/0
	VLAN	100
	DNet-2 GEM port index	2
	DNet-2 mapping index	0

Table 20-8 Data planning for service slices

Parameter	Description	Example Value
Slice ID	/	1
Slice name	/	slice-1
Slice priority	0-7	5

 Copyright © Huawei Technologies Co., Ltd.

Parameter	Description	Example Value
Queue buffer level	/	13
Bound VLAN	/	100
Number of MAC addresses	Indicates the maximum number of dynamic MAC addresses that can be learned by a dedicated network.	100
User port	/	0/3/0
Upstream port	/	0/9/1
Bandwidth of the user port in the slice instance	/	2048kbit/s
Bandwidth of the upstream port in the slice instance	/	2048kbit/s
Traffic profile	/	index 100

Procedure

Step 1 **(Optional)** Run the **slice mode { van | fans }** command to set the system to the FANS mode.

NOTE

By default, the FANS mode is used. You need to perform this step in non-FANS mode.

Step 2 Enter the board and run the **hqos mode** command to set the HQoS mode of the board to slice. You can run the **display hqos mode** command to query whether the HQoS mode of a board is slice.

```
huawei(config)#interface gpon 0/3
huawei(config-if-gpon-0/3)#hqos mode slice
huawei(config-if-gpon-0/3)#display hqos mode
huawei(config-if-gpon-0/3)#quit
huawei(config)#interface mpu 0/9
huawei(config-if-mpu-0/9)#hqos mode slice
huawei(config-if-mpu-0/9)#display hqos mode
huawei(config-if-mpu-0/9)#quit
```

Step 3 Configure a dedicated network slice.

1. Run the **dedicated-net** command to create a dedicated network slice.

   ```
   huawei(config)#dedicated-net dedicated-net-id 1 dedicated-net-name DNet-1
   huawei(config)#dedicated-net dedicated-net-id 2 dedicated-net-name DNet-2
   ```

2. Run the **dedicated-net-binding** command to add port resources for the dedicated network slice. The ports are bound to the dedicated network slice and the upstream port uses an exclusive bandwidth. PON port bandwidths are configured according to the service plan.

   – Configure resources for DNet-1.

   ```
   huawei(config-dnet-1)#dedicated-net-binding port 0/9 0 // Add upstream port resources.
     Number of ports bound successfully: 1
     Number of ports bound failed      : 0
   ```

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config-dnet-1)#dedicated-net-binding port 0/3 0 channel assure-bandwidth 700
7000  // Add service port resources and assign bandwidths.
  Number of ports bound successfully: 1
  Number of ports bound failed      : 0
```

- Configure resources for DNet-2.

```
huawei(config-dnet-2)#dedicated-net-binding port 0/9 1   // Add upstream port resources.
  Number of ports bound successfully: 1
  Number of ports bound failed      : 0

huawei(config-dnet-2)#dedicated-net-binding port 0/3 0 channel assure-bandwidth 300
3000  // Add service port resources and assign bandwidths.
  Number of ports bound successfully: 1
  Number of ports bound failed      : 0
```

3. Run the **vlan** command to create a VLAN for the dedicated network slice.

```
huawei(config)#vlan 100 dedicated-net-id 1     // Create a VLAN for DNet-1.
huawei(config)#vlan 100 dedicated-net-id 2     // Create a VLAN for DNet-2.
```

4. Add the upstream port to a VLAN.

```
huawei(config)#port vlan 100 dedicated-net-id 1 0/9 0
huawei(config)#port vlan 100 dedicated-net-id 2 0/9 1
```

5. Configure a service profile and line profile.
   - Configure a DBA profile.

```
huawei(config)#dba-profile add profile-id 31 profile-name DBA_kongzhi type3 assure
1024000 max 10000000
huawei(config)#dba-profile add profile-id 32 profile-name DBA_jiankong type4 max
10000000
```

   - Configure a service profile.

```
huawei(config)#ont-srvprofile gpon profile-id 10
huawei(config-gpon-srvprofile-10)#ont-port eth adaptive 8
huawei(config-gpon-srvprofile-10)#port vlan eth 1 transparent
huawei(config-gpon-srvprofile-10)#port vlan eth 2 transparent
huawei(config-gpon-srvprofile-10)#port vlan eth 3 transparent
huawei(config-gpon-srvprofile-10)#port vlan eth 4 transparent
huawei(config-gpon-srvprofile-10)#port vlan eth 5 transparent
huawei(config-gpon-srvprofile-10)#port vlan eth 6 transparent
huawei(config-gpon-srvprofile-10)#port vlan eth 7 transparent
huawei(config-gpon-srvprofile-10)#port vlan eth 8 transparent
huawei(config-gpon-srvprofile-10)#commit
```

   - Configure a line profile (by ONU).

```
huawei(config)#ont-lineprofile gpon profile-id 10
huawei(config-gpon-lineprofile-10)#omcc encrypt on
huawei(config-gpon-lineprofile-10)#tcont 1 dba-profile-id 31
huawei(config-gpon-lineprofile-10)#gem add 1 eth tcont 1 encrypt on
huawei(config-gpon-lineprofile-10)#gem mapping 1 0 vlan 100
huawei(config-gpon-lineprofile-10)#commit
huawei(config-gpon-lineprofile-10)#quit
```

   - Configure a line profile (by ONU Ethernet port).

```
huawei(config)#ont-lineprofile gpon profile-id 20
huawei(config-gpon-lineprofile-20)#omcc encrypt on
huawei(config-gpon-lineprofile-20)#mapping-mode port
huawei(config-gpon-lineprofile-20)#tcont 1 dba-profile-id 31
huawei(config-gpon-lineprofile-20)#tcont 2 dba-profile-id 32
huawei(config-gpon-lineprofile-20)#port-bundle eth 1 port 1,2
huawei(config-gpon-lineprofile-20)#port-bundle eth 2 port 3,4
huawei(config-gpon-lineprofile-20)#gem add 1 eth tcont 1 encrypt on
huawei(config-gpon-lineprofile-20)#gem add 2 eth tcont 2 encrypt on
huawei(config-gpon-lineprofile-20)#gem mapping 1 0 eth-bundle 1
huawei(config-gpon-lineprofile-20)#gem mapping 2 0 eth-bundle 2
huawei(config-gpon-lineprofile-20)#commit
huawei(config-gpon-lineprofile-20)#quit
```

6. Add an ONU.
   - Add ONU 0 and bind the ONU-level profile to it.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#interface gpon 0/3
huawei(config-if-gpon-0/3)#ont add 0 0 sn-auth "485754432Dxxxxxx" omci ont-lineprofile-id 10 ont-srvprofile-id 10 ont-type 10g/10g
huawei(config-if-gpon-0/3)#ont port native-vlan 0 0 eth 1 vlan 100  // Configure a default VLAN for ONT Ethernet ports
huawei(config-if-gpon-0/3)#ont port native-vlan 0 0 eth 2 vlan 100
huawei(config-if-gpon-0/3)#ont port native-vlan 0 0 eth 3 vlan 100
huawei(config-if-gpon-0/3)#ont port native-vlan 0 0 eth 4 vlan 100
huawei(config-if-gpon-0/3)#ont port native-vlan 0 0 eth 5 vlan 100
huawei(config-if-gpon-0/3)#ont port native-vlan 0 0 eth 6 vlan 100
huawei(config-if-gpon-0/3)#ont port native-vlan 0 0 eth 7 vlan 100
huawei(config-if-gpon-0/3)#ont port native-vlan 0 0 eth 8 vlan 100
```

- Add ONU 1 and bind the ONU Ethernet port granularity profile to ONU 1.

```
huawei(config-if-gpon-0/3)#ont add 0 1 sn-auth "485754432Dyyyyyy" omci ont-lineprofile-id 20 ont-srvprofile-id 10 ont-type 10g/10g
huawei(config-if-gpon-0/3)#ont port native-vlan 0 1 eth 1 vlan 100  // Configure a default VLAN for ONT Ethernet ports
huawei(config-if-gpon-0/3)#ont port native-vlan 0 1 eth 2 vlan 100
huawei(config-if-gpon-0/3)#ont port native-vlan 0 1 eth 3 vlan 100
huawei(config-if-gpon-0/3)#ont port native-vlan 0 1 eth 4 vlan 100
huawei(config-if-gpon-0/3)#ont port native-vlan 0 1 eth 5 vlan 100
huawei(config-if-gpon-0/3)#ont port native-vlan 0 1 eth 6 vlan 100
huawei(config-if-gpon-0/3)#ont port native-vlan 0 1 eth 7 vlan 100
huawei(config-if-gpon-0/3)#ont port native-vlan 0 1 eth 8 vlan 100
```

7. Run the **service-port** command to create a service port for the dedicated network.

   NOTE

   This step is mandatory when only dedicated network slices are configured. Skip this step if service slices need to be configured for a dedicated network slice.

   - Configure ONT-based service ports for DNet-1.

     ```
     huawei(config)#service-port 1000 vlan 100 dedicated-net-id 1 gpon 0/3/0 ont 0 gemport 1 multi-service user-vlan 100  // Add an ONU 0 service port and bind it to DNet-1.
     huawei(config)#service-port 1001 vlan 100 dedicated-net-id 1 gpon 0/3/0 ont 1 gemport 1 multi-service user-vlan 100  // Add a service port for ports 1–2 of ONU 1 and bind it to DNet-1.
     ```

   - Configure ONT-based service ports for DNet-2.

     ```
     huawei(config)#service-port 1002 vlan 100 dedicated-net-id 2 gpon 0/3/0 ont 1 gemport 2 multi-service user-vlan 100  // Add a service port for ports 3–4 of ONU 1 and bind it to DNet-2.
     ```

8. Configure the maximum number of learnable dynamic MAC addresses for the dedicated network.

   ```
   huawei(config)#dedicated-net dedicated-net-id 1
   huawei(config-dnet-1)#mac-address max-mac-count 100  // Configure the maximum number of learnable MAC addresses for DNet-1.

   huawei(config)#dedicated-net dedicated-net-id 2
   huawei(config-dnet-2)#mac-address max-mac-count 100  // Configure the maximum number of learnable MAC addresses for DNet-2.
   ```

9. View all created dedicated network slices.

   ```
   huawei(config)#display dedicated-net all
   { <cr>||<K> }:

   Command:
         display  dedicated-net  all
   ----------------------
   D-Net ID     Slice ID
   ----------------------
        0     -
        1     -
        2     -
   ----------------------
   Total : 3
   ```

 Copyright © Huawei Technologies Co., Ltd.

10. View details about a specified dedicated network slice.

```
huawei(config)#display dedicated-net dedicated-net-id  1  // View details about DNet-1.
{ <cr>||<K> }:

 Command:
        display dedicated-net dedicated-net-id  1
 D-Net ID    : 1
 D-Net name  : DNet-1
 Slice ID    : -
 VLAN list   : 1,100
 Description :
 Available dyn-mac number : 100
 Queue buffer class : 6
 -----------------------------------------------------
 F/ S/ P      Bandwidth(Mbps)      Ingress      Egress
 -----------------------------------------------------
 0/ 9/ 0                  -           -          -
 -----------------------------------------------------
 Network port number : 1
 ------------------------------------------------------------------------
 F/ S/ P      Bandwidth(Mbps)      Ingress      Egress    Channel ID
 ------------------------------------------------------------------------
 0/ 3/ 0                700           -          -           0
 0/ 3/ 0               7000           -          -           1
 ------------------------------------------------------------------------
 User port number : 2
 Network vport number : 0

huawei(config)#display dedicated-net dedicated-net-id  2  // View details about DNet-2.
{ <cr>||<K> }:

 Command:
        display dedicated-net dedicated-net-id  2
 D-Net ID    : 2
 D-Net name  : DNet-2
 Slice ID    : -
 VLAN list   : 1,100
 Description :
 Available dyn-mac number : 100
 Queue buffer class : 6
 -----------------------------------------------------
 F/ S/ P      Bandwidth(Mbps)      Ingress      Egress
 -----------------------------------------------------
 0/ 9/ 1                  -           -          -
 -----------------------------------------------------
 Network port number : 1
 ------------------------------------------------------------------------
 F/ S/ P      Bandwidth(Mbps)      Ingress      Egress    Channel ID
 ------------------------------------------------------------------------
 0/ 3/ 0                300           -          -           0
 0/ 3/ 0               3000           -          -           1
 ------------------------------------------------------------------------
 User port number : 2
 Network vport number : 0
```

Step 4 Create a service slice in the dedicated network slice.

1. Create a service slice under the dedicated network.

```
huawei (config) # network slice-instance slice-id 1 dedicated-net-id 1 priority 5  //Create a service slice in DNet-1.
huawei(config-slice-instance-1)#queue-buffer-class 13
```

2. Add resources to the service slice.

```
huawei(config-slice-instance-1)#nsi-binding vlan 100
huawei(config-slice-instance-1)#nsi-binding port 0/3 0 bandwidth 1024
huawei(config-slice-instance-1)#nsi-binding port 0/9 1 bandwidth 1024
huawei(config)#service-port 100 vlan 100 dedicated-net-id 1 gpon 0/3/0 ont 0 gemport 0 multi-service user-vlan 100 tag-transform translate inbound traffic-table index 100 outbound traffic-table index 100
```

 Copyright © Huawei Technologies Co., Ltd.

3. Query the service slice configuration result.

```
huawei(config)#display network slice-instance slice-id 1
{ <cr>||<K> }:

Command:
        display network slice-instance slice-id 1
Slice ID           : 3
Slice name         : slice-3
D-Net ID           : 1
Priority           : 3
Service Type       : any
Queue buffer class : 13
VLAN list          : 100
VXLAN VNI list     : -
-------------------------------------------------
F/ S/ P    Bandwidth(Mbps)    Convergence ratio
-------------------------------------------------
0/ 9/ 1              1024                1.0
-------------------------------------------------
Network port number : 1
-------------------------------------------------------------------
F/ S/ P    Bandwidth(Mbps)    Convergence ratio    Channel ID
-------------------------------------------------------------------
0/ 3/ 0              1024                1.0            -
-------------------------------------------------------------------
User port number : 1
Network vport number : 0
```

----End

20.4.2 Service Slice Deployment

This section uses the HD video golden slice as an example to describe how to create a slice instance and provision services.

20.4.2.1 Service Slice Deployment Example (NCE)

Prerequisites

- OLT model: MA5800-X17, X15, X7, and X2
- OLT version: V100R021C00 or later
- NCE-FAN version: V100R021C00 or later
- To use this feature, you need to purchase and install the *Access Network Auto Configuration Package License* and *Access Service Slice License*.

  NOTE

  For details about how to load a license, see **NCE-FAN Application Operation and Maintenance > Basic Applications > System Settings** > **License Management** in the **iMaster NCE-FAN Product Documentation**.

  Carrier link: **License Management**

  Enterprise network link: **License Management**
- The NEs and NMS have been installed and are running properly. The network is running properly. A management channel has been established.
- The app to be guaranteed and the corresponding server IP address list are available.
- Slice instance parameters have been planned.

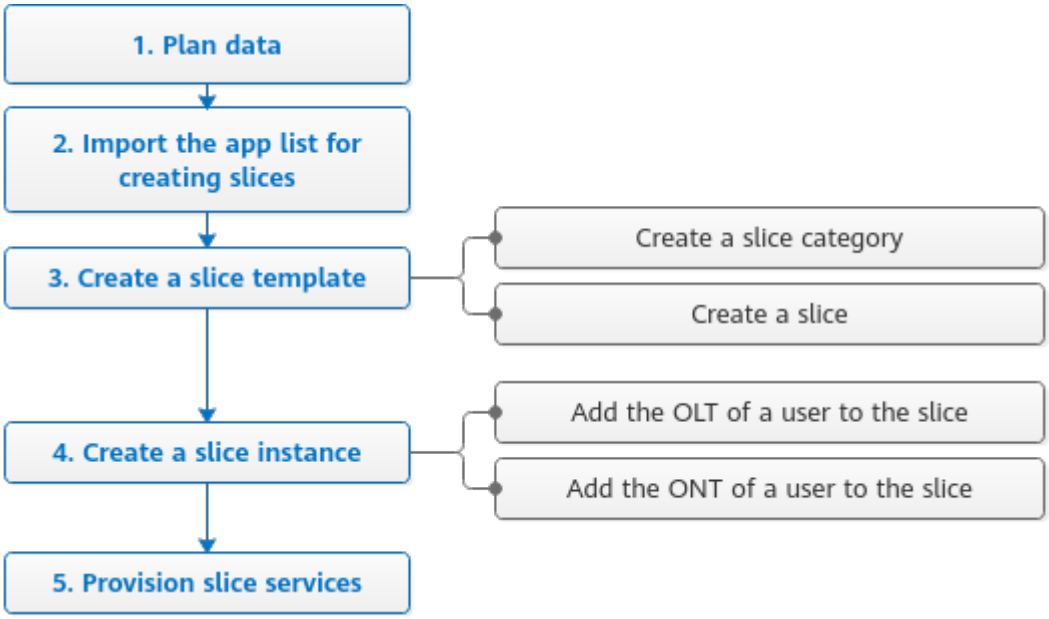
 Copyright © Huawei Technologies Co., Ltd.

NOTE

When adding slices in the reconstruction scenario, ensure that:

- A GEM port has been specified for the service port configured on the ONT. If the GEM port is not specified, contact Huawei engineers to modify the configuration.
- The ONT type is not auto.

Slice Configuration Process



Data Planning Example

Table 20-9 Key parameter planning for slicing

Category	Parameter	Description	Example Value
Slice template	Create Slice Category	Customized	Video
	Slice Name	Customized	SliceName1
	SLA template (different templates have different parameters)	Preset and directly selected	Video HD
	Network-side Settings	Preset and directly selected	VLAN
	Outer VLAN ID	Customized	100
	User-side Settings	Preset and directly selected	VLAN
	Outer VLAN ID	Customized	1001

 Copyright © Huawei Technologies Co., Ltd.

Category	Parameter	Description	Example Value
Slice instance NE information	Slice Name	Customized	OLT
	NE IP Address	Customized	10.xxxxxx
	Uplink Port	Customized	0/9/0
	Uplink Bandwidth	Customized	1000Mbit/s
	PON Port	Customized	0/3/0
	Downlink Bandwidth	Customized	1000Mbit/s
	ONT SN	Customized	48575443ACC14071

Procedure

Step 1 Import the app and the corresponding server IP address list into NCE for binding a slice instance.

1. On the NCE home page, click the **Access Controller** app and choose **Network Slice** from the main menu.

Figure 20-8 Home page

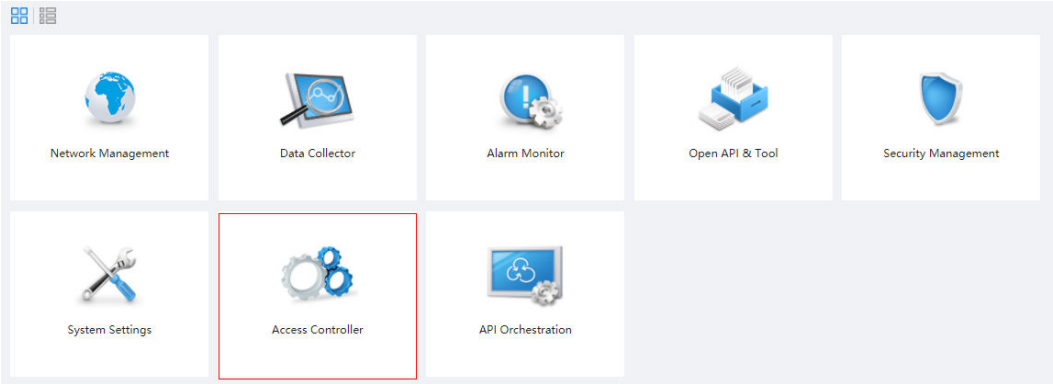
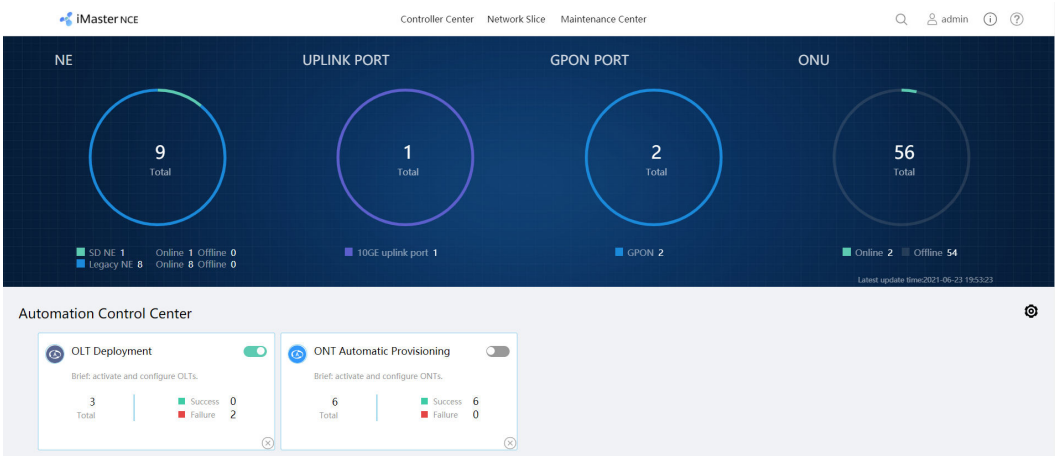


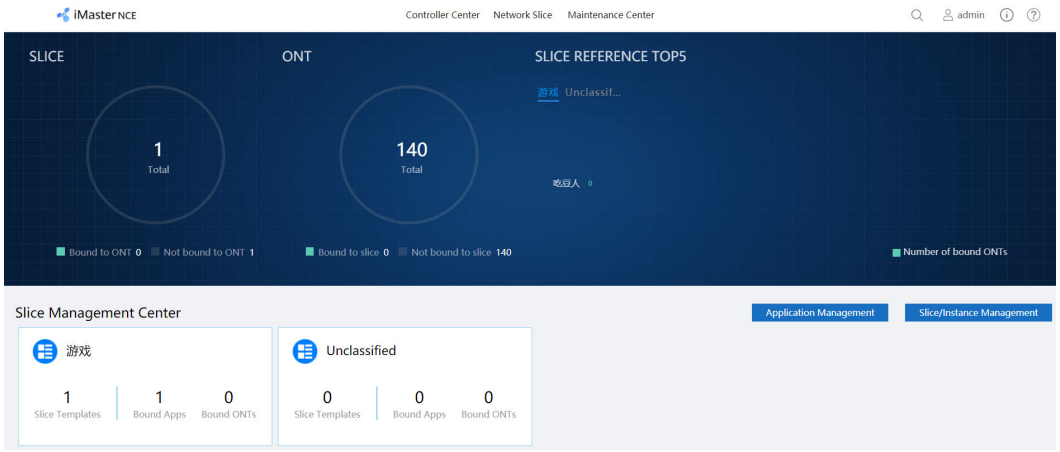
Figure 20-9 Access Controller page



 Copyright © Huawei Technologies Co., Ltd.

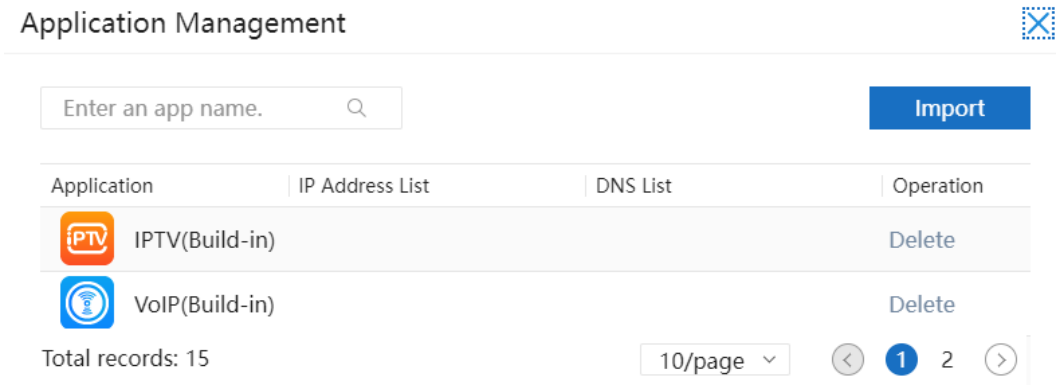
2. In the **Slice Management Center** area, click **Application Management**.

Figure 20-10 Network slice/slice management center



3. Click **Import**. In the displayed dialog box, download a template. Enter application information on the local PC and upload the application template file (JSON file) and application icon package (ZIP package) to NCE.

Figure 20-11 Importing an app



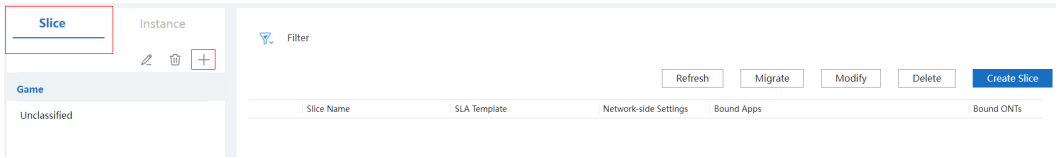
NOTE

For details about the requirements on the JSON file and ZIP package, see the online help on the NCE import page.

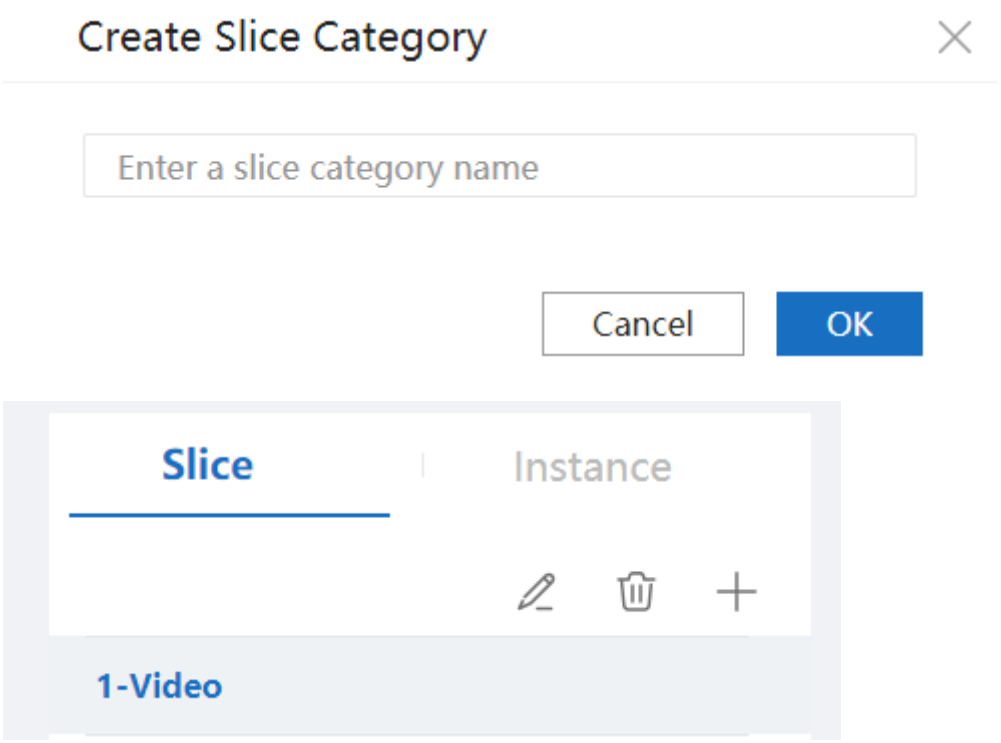
Step 2 Create a slice category.

1. In the **Slice Management Center** area, click **Slice/Instance Management**.
2. On the **Slice** tab page, click +. In the dialog box that is displayed, enter the slice category name and click **OK**.

Figure 20-12 Slice/instance management page



 Copyright © Huawei Technologies Co., Ltd.



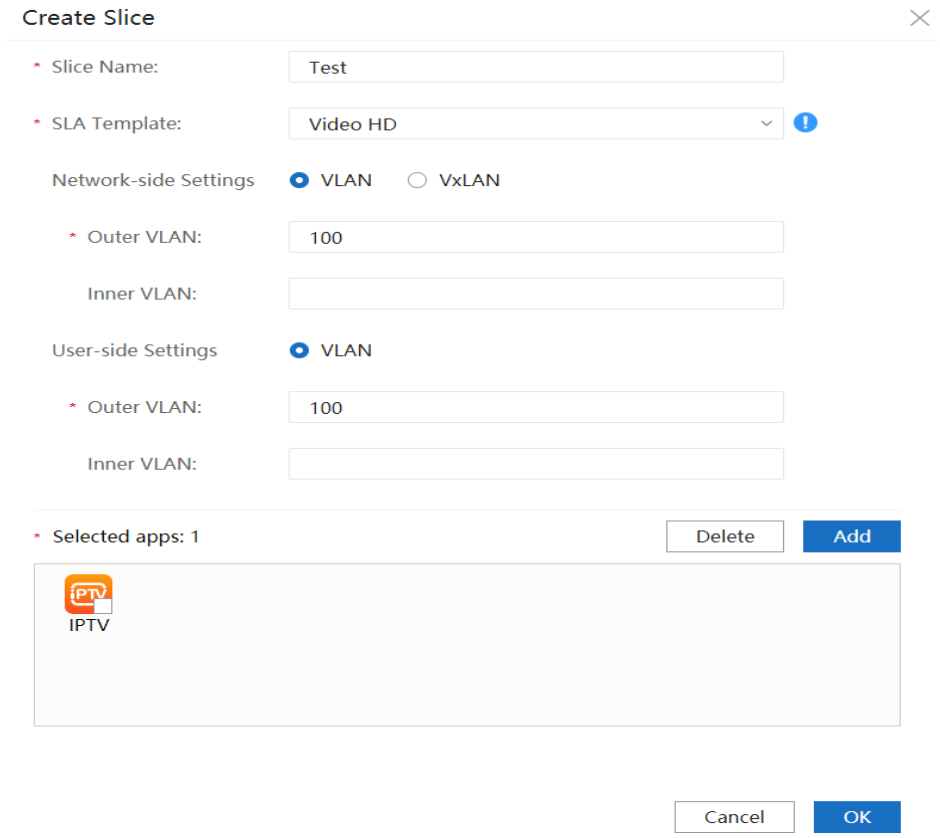
NOTE

NCE supports a maximum of eight slice categories.

Step 3 Create a slice.

1. Select a created slice category and click **Create Slice** in the right pane.
2. In the **Create Slice** dialog box, set related parameters, select the app to be bound, and click **OK**.

 Copyright © Huawei Technologies Co., Ltd.



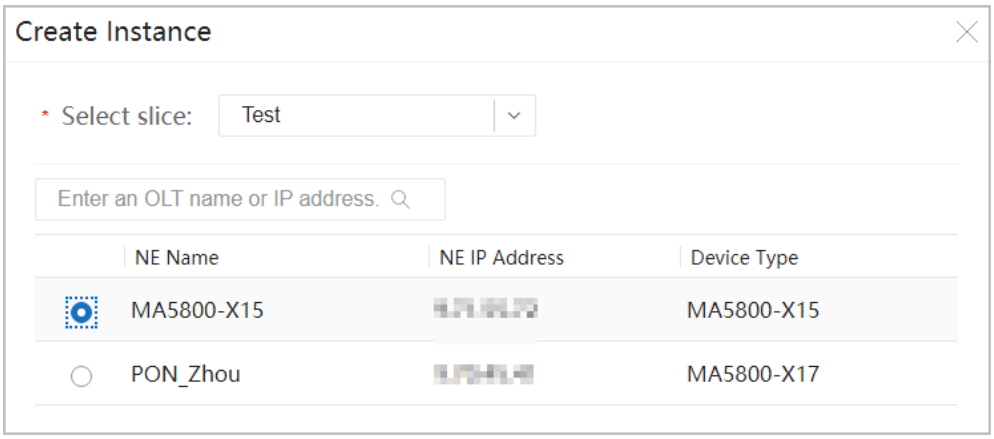
NOTE

- NCE supports a maximum of 1024 slices.
- One slice can be bound to a maximum of 16 apps.
- After selecting an SLA template, you can click to view the parameter values of the template.
- For common services, such as the Internet access service, a VLAN can be configured on the network side to use the VLAN tunnel.
- For services that have high requirements on bandwidth and latency, such as education services, VXLAN can be configured on the network side to use the VXLAN tunnel.

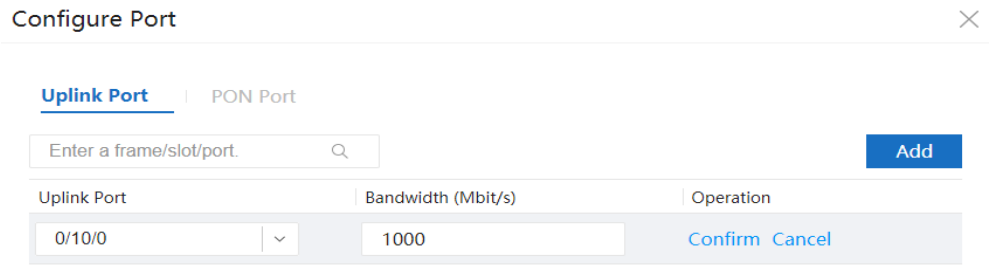
Step 4 Create a slice instance.

1. **Add an OLT to the slice instance.** Click the **Instance** tab, select a slice category on the left, and click **Create Instance** on the right. In the **Create Instance** dialog box that is displayed, select the slice and OLT, and click **OK**.

 Copyright © Huawei Technologies Co., Ltd.



2. **Configure detailed OLT parameters for the slice instance.** Click **Configure Port** in the **Operation** column. In the dialog box that is displayed, configure the uplink port and PON port.
   a. On the **Uplink Port** tab page, click **Add**, configure the uplink port and uplink bandwidth, and click **OK**.
   b. On the **PON Port** tab page, click **Add**, configure the PON port and downlink bandwidth, and click **OK**.

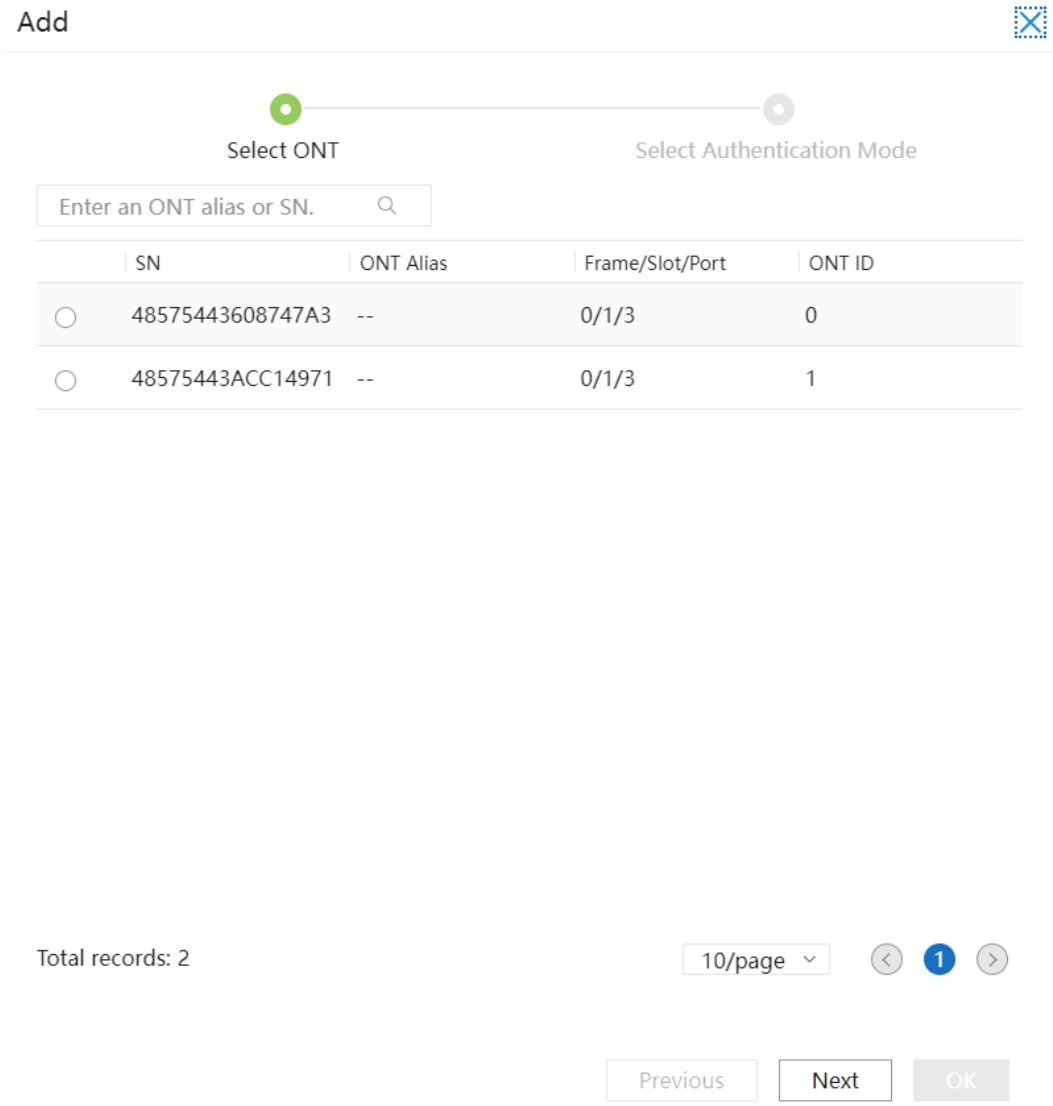


Step 5 Bind an ONT and provision the slice service.

1. Click **Configure ONT** in the **Operation** column. In the dialog box that is displayed, select an ONT.

2. Click **Add**. In the dialog box that is displayed, select an ONT and click **Next**.

 Copyright © Huawei Technologies Co., Ltd.



a. Select the ONT authentication mode and click **OK**.

b. After the ONT passes the authentication, NCE delivers slice configurations to the ONT.

Step 6 **View the configuration result**. In the Slice Management Center, view information about the added slices, bound applications, and ONTs in different slices.

Step 7 **A subscriber uses services or apps in the slice.**

----End

20.4.2.2 Service Slice Deployment Example (CLI)

Prerequisites

- OLT model: MA5800-X17, X15, X7, and X2
- OLT version: V100R021C00 or later
- The NEs have been installed and are running properly. The network is running properly. A management channel has been established.

 Copyright © Huawei Technologies Co., Ltd.

- An app and its server IP address list have been imported into to NCE. For details, see **20.4.2.1 Service Slice Deployment Example (NCE)** > **Step 1**.
- Slice instance parameters have been planned.
- A traffic profile with the same CIR and PIR has been configured.

Data Planning Example

Table 20-10 Key data planning

Parameter	Example Value
Slice ID	3
Slice name	slice-3
Slice priority	3
Queue buffer level	13
Bound VLAN	100
User port	0/3/0
Uplink port	0/9/1
Bandwidth of the user port in the slice instance	2048kbit/s
Bandwidth of the uplink port in the slice instance	2048kbit/s
Traffic profile	index 100

Procedure

Step 1 **(Optional) Set the system to the FANS mode.** Run the **slice mode { van | fans }** command to set the system to the FANS mode.

NOTE

By default, the FANS mode is used. You need to perform this step in non-FANS mode.

Step 2 **Create a slice instance.** Run the **network slice-instance** command to create a slice and set a priority.

NOTE

The priority determines whether the slice bandwidth can be preempted. The value 0 indicates a copper slice (the default priority is 0 and no slice needs to be created), values 1 to 5 indicate silver slices, and values 6 to 7 indicate golden slices. Golden slices have the highest priority.

```
huawei(config)#network slice-instance slice-id 3 slice-name slice-3 priority 3 service-type any
```

Step 3 **Configure slice instance parameters.**

1. Run the **queue-buffer-class** command to set the buffer level of a slice queue.

```
huawei(config-slice-instance-3)#queue-buffer-class 13
```

 Copyright © Huawei Technologies Co., Ltd.

2. Run the **nsi-binding vlan** command to bind a VLAN to a slice instance.
```
huawei(config-slice-instance-3)#nsi-binding vlan 100
```
3. Enter the board and run the **hqos mode** command to set the HQoS mode of the board to slice. You can run the **display hqos mode** command to query whether the HQoS mode of a board is slice.
```
huawei(config)#interface gpon 0/3
huawei(config-if-gpon-0/3)#hqos mode slice
huawei(config)#interface mpu 0/9
huawei(config-if-mpu-0/9)#hqos mode slice
```
4. Run the **nsi-binding port** command to bind a user port and uplink port to the slice instance and set their bandwidths in the current slice instance.
```
huawei(config-slice-instance-3)#nsi-binding port 0/3 0 bandwidth 1024
huawei(config-slice-instance-3)#nsi-binding port 0/9 1 bandwidth 1024
```
5. Run the **service-port** command to configure a service port.
```
huawei(config)#service-port 100 vlan 100 gpon 0/3/0 ont 0 gemport 0 multi-service user-vlan 100 tag-transform translate inbound traffic-table index 100 outbound traffic-table index 100
```

Step 4 Query the configuration results.
```
huawei(config)#display network slice-instance all
{ <cr>||<K> }:
  Command:
          display network slice-instance all
  ------------------------------------------------------------
  Slice ID    Priority    Queue buffer class    Service type
  ------------------------------------------------------------
       3          3                   13     any
  ------------------------------------------------------------
  Total : 1
huawei(config)#display network slice-instance slice-id 3
{ <cr>||<K> }:

  Command:
         display network slice-instance slice-id 3
  Slice ID           : 3
  Slice name         : slice-3
  Priority           : 3
  Service Type       : any
  Queue buffer class : 13
  VLAN list          : 100
  VXLAN VNI list     : -
  -------------------------------------------------
  F/ S/ P    Bandwidth(Mbps)    Convergence ratio
  -------------------------------------------------
  0/ 9/ 1             1024                1.0
  -------------------------------------------------
  Network port number : 1
  ---------------------------------------------------------------
  F/ S/ P    Bandwidth(Mbps)    Convergence ratio    Channel ID
  ---------------------------------------------------------------
  0/ 3/ 0             1024                1.0            -
  ---------------------------------------------------------------
  User port number : 1
  Network vport number : 0
```

----End

20.5 Reference

 Copyright © Huawei Technologies Co., Ltd.

20.5.1 Parameter Reference for a Service Slice SLA Template

Template Name	Parameter	Value
HD education live broadcast	Uplink bandwidth	20000000bit/s
	Downlink bandwidth	20000000bit/s
	Maximum latency	60ms
	Maximum jitter	10ms
	Maximum packet loss rate	0.01%
4K education live broadcast	Uplink bandwidth	50000000bit/s
	Downlink bandwidth	50000000bit/s
	Maximum latency	60ms
	Maximum jitter	10ms
	Maximum packet loss rate	0.01%
8K education live broadcast	Uplink bandwidth	200000000bit/s
	Downlink bandwidth	200000000bit/s
	Maximum latency	60ms
	Maximum jitter	10ms
	Maximum packet loss rate	0.01%
HD video	Uplink bandwidth	20000000bit/s
	Downlink bandwidth	20000000bit/s
	Maximum latency	60ms
	Maximum jitter	10ms
	Maximum packet loss rate	0.01%
4K video	Uplink bandwidth	50000000bit/s
	Downlink bandwidth	50000000bit/s
	Maximum latency	60ms
	Maximum jitter	10ms
	Maximum packet loss rate	0.01%
8K video	Uplink bandwidth	200000000bit/s
	Downlink bandwidth	200000000bit/s
	Maximum latency	60ms
	Maximum jitter	10ms

 Copyright © Huawei Technologies Co., Ltd.

Template Name	Parameter	Value
	Maximum packet loss rate	0.01%
Weak-interaction VR	Uplink bandwidth	1024bit/s
	Downlink bandwidth	1024bit/s
	Maximum latency	10ms
	Maximum jitter	4ms
	Maximum packet loss rate	0.01%
Strong-interaction VR	Uplink bandwidth	200000000bit/s
	Downlink bandwidth	200000000bit/s
	Maximum latency	10ms
	Maximum jitter	4ms
	Maximum packet loss rate	0.01%
PC games	Uplink bandwidth	20000000bit/s
	Downlink bandwidth	20000000bit/s
	Maximum latency	10ms
	Maximum jitter	5ms
	Maximum packet loss rate	0.01%
HD cloud gaming	Uplink bandwidth	20000000bit/s
	Downlink bandwidth	20000000bit/s
	Maximum latency	10ms
	Maximum jitter	4ms
	Maximum packet loss rate	0.01%
4K cloud gaming	Uplink bandwidth	50000000bit/s
	Downlink bandwidth	50000000bit/s
	Maximum latency	10ms
	Maximum jitter	4ms
	Maximum packet loss rate	0.01%
8K cloud gaming	Uplink bandwidth	200000000bit/s
	Downlink bandwidth	200000000bit/s
	Maximum latency	10ms
	Maximum jitter	4ms

 Copyright © Huawei Technologies Co., Ltd.

Template Name	Parameter	Value
	Maximum packet loss rate	0.01%
Multi-party voice conference	Uplink bandwidth	2048bit/s
	Downlink bandwidth	2048bit/s
	Maximum latency	20ms
	Maximum jitter	6ms
	Maximum packet loss rate	0.01%

20.5.2 Acronyms, Abbreviations, and Basic Concepts

Abbreviation	Full Spelling
MANO	Management and Orchestration
NF	Network Function
NFV	Network Functions Virtualisation
NSI	Network Slice Instance
NSMF	Network Slice Management Function
NST	Network Slice Template
NSST	Network Slice Subnet Template
NSSI	Network Slice Subnet Instance
NSSMF	Network Slice Subnet Management Function
CSMF	Communication Service Management Function
VNF	Virtualized Network Function
NCE	Network Cloud Engine
OLT	Optical line terminal
FTTH	Fiber to the home
SST	Servcie and slice type
NSI	Network slice instance
FAN	Fixed Access network
GPON	Gigabit-capable Passive Optical Networks
VAN	Virtual access node

 Copyright © Huawei Technologies Co., Ltd.

Abbreviation	Full Spelling
VXLAN	Virtual Extensible LAN
OSS	Operations support system
CP	Content provider
BNG	Broadband network gateway
SR	Service router
D-NET	Dedicate Network
ISP	Internet Service Provider
VNO	Virtual Network Operator

20.5.3 Reference Standards and Protocols for Slicing

Number	Description
TR370	Fixed Access Network Sharing

 Copyright © Huawei Technologies Co., Ltd.

21 Aggregation Management

21.1 Changes in Aggregation Management
21.2 Hardware Version Description for Aggregation Management
21.3 Introduction to the Aggregation Management Solution
21.4 Basic Concepts of Aggregation Management
21.5 Implementation Principle of Aggregation Management
21.6 Application Scenario of Aggregation Management
21.7 Aggregation Management Configuration

21.1 Changes in Aggregation Management

Product Version	Change Description
V100R022C0x	First version to support this feature.
V100R022C1x	The MA5800-X2 can be used as a slave OLT.
V100R023C0x	Modified the slave OLT self-healing function.

21.2 Hardware Version Description for Aggregation Management

Hardware of Master OLT	Hardware of Slave OLT	Required Version
MA5800-X17, MA5800-X15, MA5800-X7	MA5801 series and MA5801S series	V100R022C0x and later

 Copyright © Huawei Technologies Co., Ltd.

Hardware of Master OLT	Hardware of Slave OLT	Required Version
MA5800-X17, MA5800-X15, MA5800-X7	MA5800-X2	V100R022C1x and later

NOTE

- The versions of the master and slave OLTs must match. For details about the version mapping, see the MA5800 Release Notes.
- The fiber distance between the master OLT and the slave OLT is only restricted by the optical module of the master OLT. For details about the transmission distance of an optical module, see MA5800 Optical Module Description.
- After aggregation management is configured successfully, the master and slave OLTs cannot be downgraded to versions earlier than V100R022C0x. If downgraded, the master and slave OLTs do not support aggregated management. As a result, the slave OLTs cannot be managed by the master OLT and services are interrupted. When the MA5800-X2 is used as a slave OLT, the slave OLT cannot be downgraded to a version earlier than V100R022C1x.
- When the MA5800-X2 is used as a slave OLT, the control board of the slave OLT can be only H901MPSG-G or H901MPSD-G, and only the port on the control board can be used as an aggregation management port. The port on an upstream board cannot be used for aggregation management. In addition, it is recommended that active/standby boards be connected to the master OLT in dual-fiber aggregation mode. Otherwise, in single-fiber scenarios, a single control board fault on a slave OLT may cause link disconnection and service interruption.
- Slave OLTs with factory versions of V100R018SPC600 and earlier, V100R019C11SPC300 and earlier patch versions, and V100R021C11SPC200 and earlier patch versions do not support aggregation management. For details about the factory versions, contact local Huawei technical support.

21.3 Introduction to the Aggregation Management Solution

Background

With extensive deployment of all-optical networks, optical line terminals (OLTs) will become more diversified (such as Blade OLTs and Pizza OLTs) and scenario-based. If OLTs are still separately managed by the NMS, the management scale will multiply, and network deployment, upgrade, and maintenance will become more complex.

In this context, Huawei launches the aggregation management solution to simplify network management and device deployment, and reduce the operational expenditure (OPEX) and total cost of ownership (TCO) for customers.

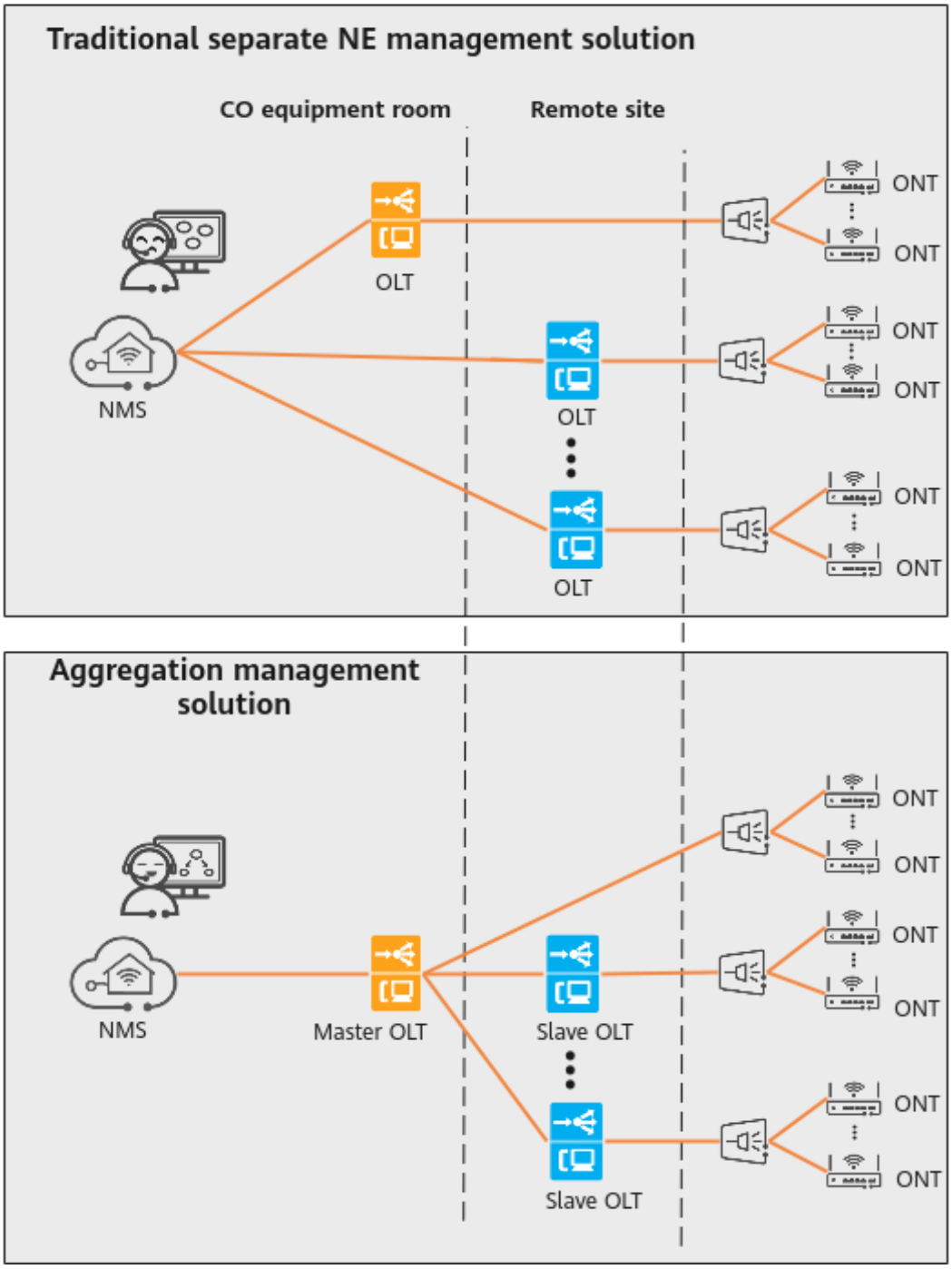
Solution Introduction

In aggregation management, a master OLT with complex functions is deployed in the CO equipment room, and the functions of remote slave OLTs are simplified. In

 Copyright © Huawei Technologies Co., Ltd.

this way, the device management and deployment are simplified. You only need to manage the master OLT in the CO equipment room on the NMS to manage all OLTs on the entire network.

Figure 21-1 Comparison between the aggregation management solution and traditional separate NE management solution



Solution Benefits

Simplified network management

 Copyright © Huawei Technologies Co., Ltd.

- All remote slave OLTs are connected to the master OLT in the CO equipment room. The NMS manages OLTs on the entire network as a single NE and issues data to the master OLT. Then, the master OLT issues the data to slave OLTs.
- On slave OLTs, service provisioning interfaces remain unchanged, and the user interfaces are the same as those for a master OLT service board, ensuring consistent user experience.

Reducing device deployment and O&M costs

- Deploying a new remote slave OLT is equivalent to adding a master OLT service board, and no re-interconnection with the upper-layer OSS or NMS is required.
- Software commissioning is not required for a remote slave OLT. You only need to visit the installation site once and power on a slave OLT. Services are automatically provisioned and deployed.
- When upgrading or replacing a remote slave OLT, you do not need to manually load the configurations of the remote slave OLT. The master OLT automatically issues the configurations to the remote slave OLT.
- A maximum of 8 remote slave OLTs can be connected.

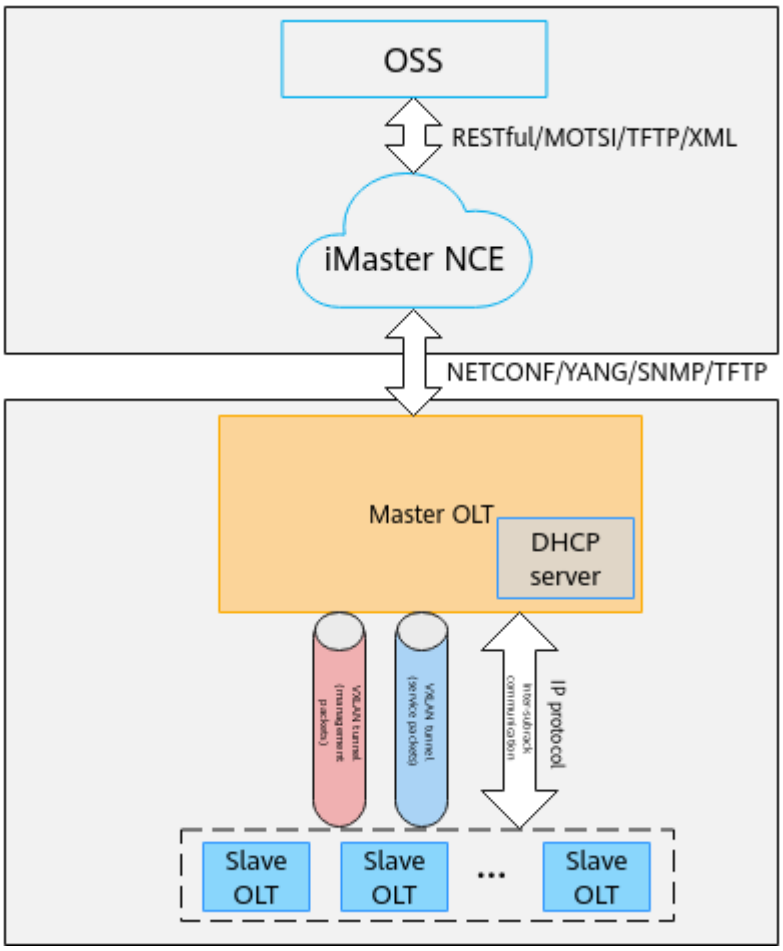
21.4 Basic Concepts of Aggregation Management

An aggregation management system is a highly autonomous small network. Externally, only the master OLT in a CO equipment room needs to be managed to manage OLTs on an entire network. A series of protocols must be supported internally, and system security, isolation, and reliability must be considered.

Figure 21-2 shows the aggregation management system model.

 Copyright © Huawei Technologies Co., Ltd.

Figure 21-2 Aggregation management system model



VXLAN Tunnel

The most significant feature of aggregation management is that a closed autonomous network (underlay network) is automatically created between the master and slave OLTs. The routing domain of the network is limited to only between the master and slave OLTs. The network serves only to connect the master and slave OLTs, and does not involve external interoperation.

In addition, VXLAN tunnels are constructed for the network. All subsequent management and service packets between the master and slave OLTs are carried in different VXLAN tunnels for information exchange and transmission between the master and slave OLTs. A VXLAN tunnel is equivalent to an inter-board communication channel in a single-subrack system.

NOTE

The underlay network can be a Layer 2 or Layer 3 network. V100R022C0x supports only Layer 2 networks. The master and slave OLTs are directly connected in the same IP network segment. The IP channels between underlay networks are interconnected through ARP learning.

 Copyright © Huawei Technologies Co., Ltd.

Aggregation Management Network

An aggregation management network is located at the upper layer of the OLTs, usually on the metro network or SR/BNG. The management data is transmitted from the upstream port of the master OLT to the management network.

- Because the master and slave OLTs are managed in aggregated mode, protocols such as SNMP, CLI, and NETCONF/YANG of the NMS are terminated on the master OLT.
- The management protocol between the master and slave OLTs is the inter-board communication protocol which is an overlay service carried on the underlay network and requires an independent bridge domain (BD).
- An overlay management network is also a Layer 2 network, which uses a reserved VLAN of the system. Similar to the VLAN of the underlay network, the VLAN of the overlay network is an S-VLAN, which occupies VLAN specifications of the system.

Aggregation Management Channel

After a VXLAN tunnel is created successfully, the IP channel between the master and slave OLTs is set up. Then, a management channel needs to be created.

- The management channel is the key to maintain the communication between the master and slave OLTs. The management IP address is used for inter-subrack communication.
- The management IP address uniquely identifies the master or slave OLTs. Packets can be constructed and sent after the destination IP address (peer management IP address) is obtained.
- The management VLAN is a reserved VLAN of the system, and a Layer 3 interface can be created for a reserved VLAN.

  NOTE

  - A reserved VLAN is a VLAN that is reserved in the system to realize potential functions.
  - The initial IDs of reserved VLANs of the master OLT cannot be modified.

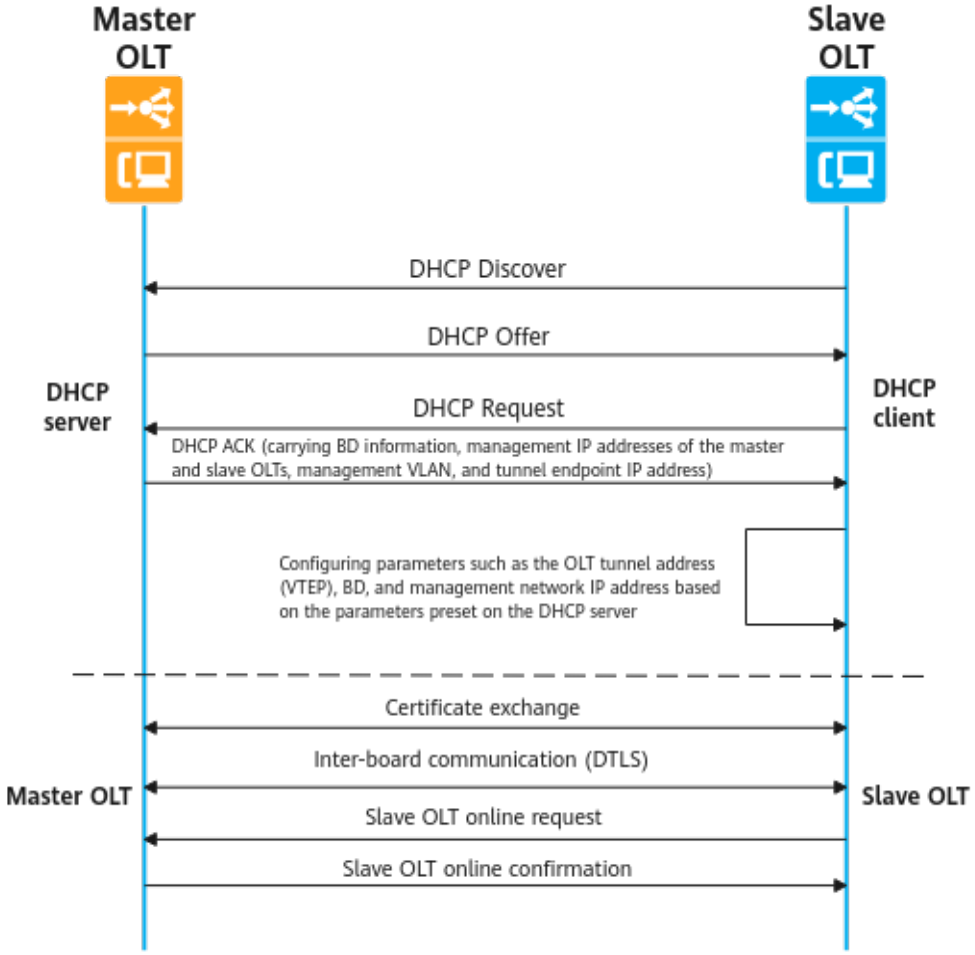
21.5 Implementation Principle of Aggregation Management

In the aggregation management solution, the master OLT and multiple slave OLTs form a network, so that all OLTs are regarded as a single NE on the NMS. Slave OLTs can be automatically deployed and replaced, simplifying network management and enabling elastic capacity expansion.

 Copyright © Huawei Technologies Co., Ltd.

Automatic Deployment

Figure 21-3 Automatic deployment process for a remote slave OLT



1. A slave OLT obtains the tunnel IP address through DHCP, and obtains the BD configurations, as well as the ID and management IP address of the slave OLT. The DHCP parameters should include the link aggregation configurations of the upstream ports.

   NOTE

   Currently, only one DHCP server can be used for aggregation management. The VLANs of the DHCP server can only be VLANs 1 to 16.

2. Based on the information obtained, the system configures the local tunnel, BD, and IP address of the slave OLT.
3. The system initiates slave OLT registration. Certificate authentication is supported for registration, and DTLS authentication is supported for the inter-board communication protocol.

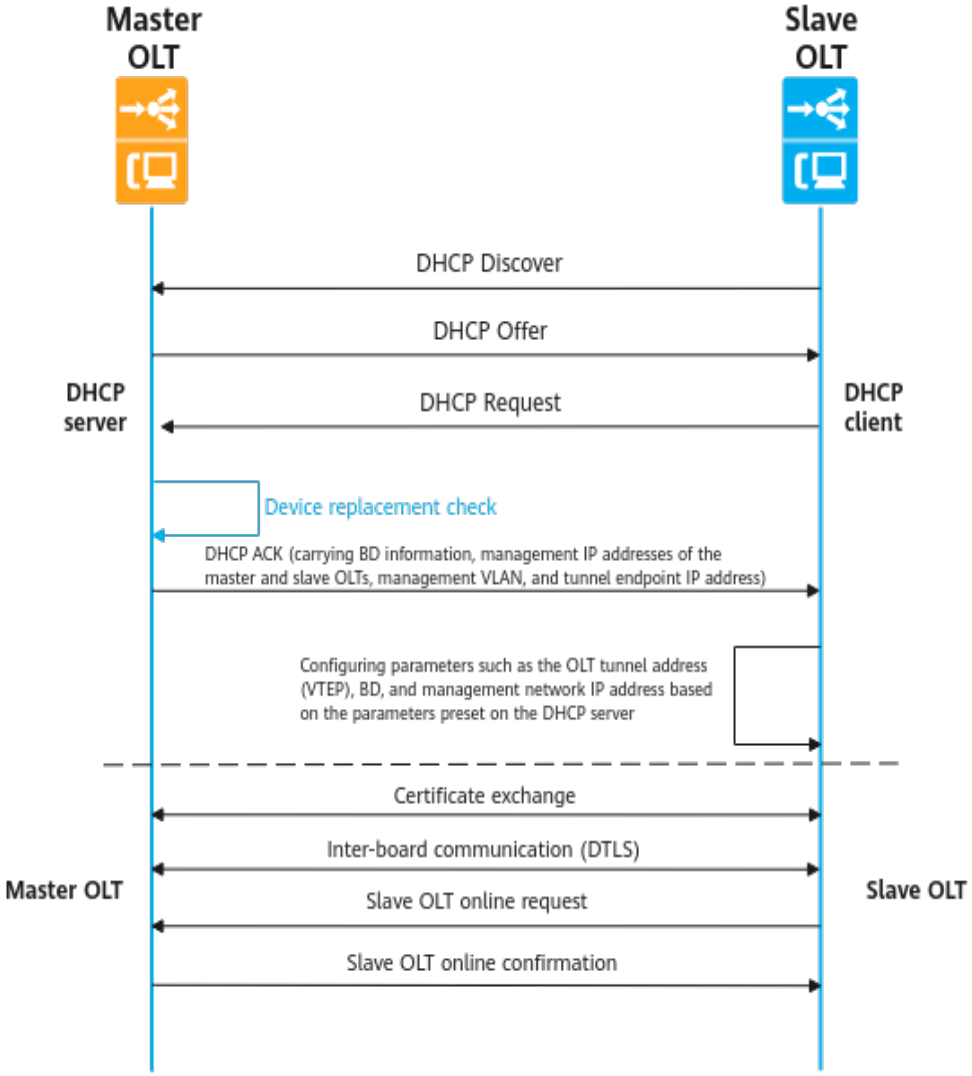
 Copyright © Huawei Technologies Co., Ltd.

NOTE

- When a slave OLT goes online in independent mode, DHCP configurations are required again.
- When a slave OLT goes online for the first time, a mode switching process is triggered, and the OLT restarts one more time. In this process, the master OLT reports a slave OLT auto-find and auto-disappear event.
- It takes 25 to 30 minutes for a slave OLT to go online for the first time. When going online for the first time, a slave OLT needs to discover the master OLT by port and VLAN polling. It is recommended that you plan a management VLAN with a smaller ID and use a port with a smaller ID to connect to a slave OLT so that a minimum time is required for going online.

Automatic Replacement

Figure 21-4 Automatic replacement process for a remote slave OLT



The automatic replacement process is similar to the automatic deployment process. The main difference is that automatic replacement requires device replacement check. The ESNs of the old and new slave OLTs are input to the

 Copyright © Huawei Technologies Co., Ltd.

device management module of the aggregation management system in advance so that the master OLT establishes the replacement relationship between the old and new slave OLTs. In this way, when the new slave OLT goes online, the configuration information of the old slave OLT is reused, that is, the new slave OLT inherits the ID of the old slave OLT.

NOTICE

Currently, a slave OLT can be replaced only with a slave OLT of the same model.

Self-healing for a Slave OLT

If the link between the master and slave OLTs is disconnected for more than 30s, the slave OLT resets for self-healing. If the link fault is rectified after the reset, the slave OLT goes online again. If a slave OLT needs to be restored to the independent mode, you need to manually switch the mode.

NOTICE

- In V100R022C1xSPH207 and earlier versions, if the link fault is not rectified within 15 minutes after a slave OLT is reset, the slave OLT switches back to the independent mode.
- In V100R022C1xSPH207 and later versions, if you want to switch from the slave mode to the independent mode, run the **Reboot system Independent mode** command in the local BIOS. If a slave OLT has two control boards, remove the standby control board, switch the active control board to the independent mode, and then insert the standby control board. The standby control board automatically synchronizes the mode of the active control board. The preset password for logging in to the local BIOS of a slave OLT is huawei123.

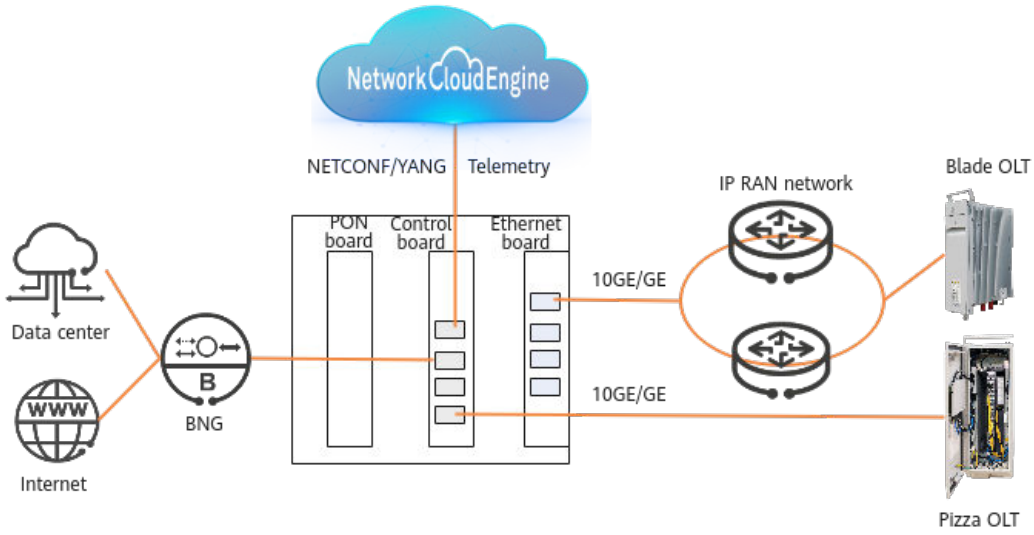
21.6 Application Scenario of Aggregation Management

Aggregation management is mainly used when multiple slave OLTs are deployed in scattered mode. The aggregation management port of the master OLT can be the upstream port on the control board or the 10GE/GE port on the Ethernet interface board.

In the OLT aggregation management system, a slave OLT (Pizza OLT or Blade OLT) can be connected to the master OLT in the CO equipment room in the following two modes:

- Connected through the IP-RAN backhaul network
- Directly connected across a Layer 2 LAN

 Copyright © Huawei Technologies Co., Ltd.



21.7 Aggregation Management Configuration

21.7.1 Digital Certificate Loading Process

In the aggregation management security channel scenario, you need to load certificates to master subracks for V100R022C00–V100R022C10 versions. Perform the following steps to load the certificates:

NOTE

In the aggregation management scenario, the CA certificates and CRL corresponding to the Huawei-initial identity certificates can be loaded using SFTP, but the identity certificates used in the aggregation management scenario cannot be replaced.

1. You can run the **display pki certificate trusted-ca list board** *frameid slotid* command to check whether the master subrack has built-in huawei_equipment_root_ca.crt and huawei_access_product_ca.crt certificates. If yes, you do not need to reload the certificates. If no, perform the following steps to load the certificates to the active and standby control boards in the master subrack.
2. Visit **https://support.huawei.com/pki/pkidetail** and download the root certificate file Huawei Equipment Root CA.pem and the level-2 CA certificate file Huawei Fixed Network Product CA.pem.
3. Run the **ssh sftp set** command to set the user name and password of the SFTP server.
4. Run the **load pki certificate trusted-ca pem sftp** [**ipv6**] *serverIpAddress filename* **board** *frameid slotid* command to load the downloaded CA certificates to the active and standby control boards in the master subrack.
5. Run the **display pki certificate trusted-ca list board** *frameid slotid* command to check whether the certificates are successfully loaded to the certificate domain for aggregation management.

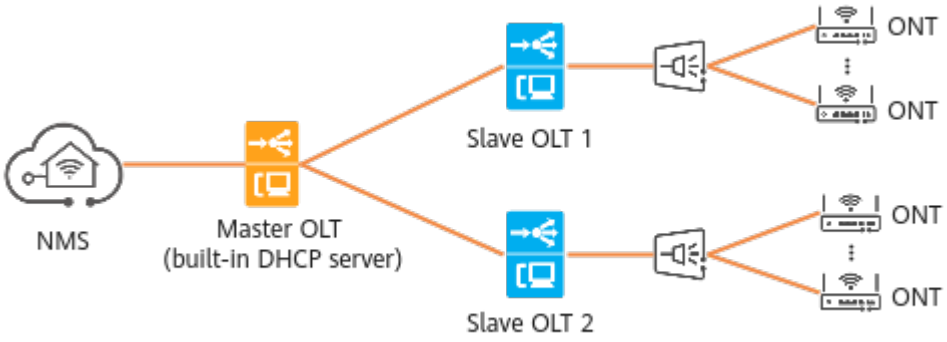
In the aggregation management security channel scenario, you do not need to load certificates to master subracks for V100R023C00 and later versions

 Copyright © Huawei Technologies Co., Ltd.

Huawei root certificates huawei_equipment_root_ca.crt and huawei_access_product_newroot.crt are initial in a software package and are automatically imported to the certificate list for aggregation management, so that the master subracks can authenticate the slave subracks. Run the **display pki certificate trusted-ca list board** *frameid slotid* command to query the certificates.

21.7.2 Typical Networking and Configuration Process

Typical Networking



Prerequisites

- A slave OLT must support certificate pre-configuration. The master OLT must support certificate pre-configuration or manual digital certificate loading. For details about how to query and load a digital certificate, see **1.7.1 Process of Loading a Digital Certificate**.
- The software versions of the master and slave OLTs support the aggregation management feature.
- A slave OLT must be started with an empty database and cannot be logged in after the startup. Otherwise, the slave OLT will be restored to the independent mode.
- It is recommended that aggregation management be used with NCE of V100R022C1x and later versions. Currently, third-party OSSs and U2000 do not fully support aggregation management.
- Do not use the centralized management and slicing features at the same time on the master OLT. Otherwise, services on the slave OLTs will be abnormal.
- When configuring aggregation management on the master OLT, ensure that DCN is disabled globally. You can run the **display dcn global** command to check whether DCN is enabled globally. If DCN is enabled globally, run the **undo dcn** command to disable it. If STP has been enabled globally, disable STP on the port connected to a slave OLT to prevent intermittent link disconnection caused by STP calculation. You can run the **stp port disable** command to disable STP.

 Copyright © Huawei Technologies Co., Ltd.

Configuration Procedure

NOTE

- In versions earlier than V100R023C0x, the aggregation group of the ports connecting the master and slave OLTs supports only the manual mode. You can run the **display link-aggregation all** command to check the working mode of the aggregation group on the master OLT. If the working mode is set to LACP, you can run the **undo link-aggregation** command to delete the working mode of the aggregation group and reconfigure the working mode.
- In V100R023C1x and later versions, aggregation management supports ARP attack defense and MAC anti-flapping for slave OLTs. After a slave OLT goes online, the master OLT generates a static ARP entry and static MAC address based on the aggregation management VLAN and saves them to a configuration file. If you run the **active configuration system** command to downgrade the master OLT to an earlier version, residual static ARP entries and static MAC addresses exist. As a result, a slave OLT may fail to go online. You can run the **undo arp** and **undo mac-address static** commands to manually delete the static ARP entries and MAC addresses generated by the master OLT based on the aggregation management VLAN so that the slave OLTs can go online successfully.
- When an aggregation network is configured for the master OLT, it is recommended that you run the **ring check enable** and **ring check role-type detection enable** commands to enable ring network detection on the master OLT; otherwise, looped ports cannot be detected when the master and slave OLTs form a ring network. If ring network detection is enabled, a temporary ring network occurs. In this case, the port cascaded to the master OLT may report a ring network alarm or event which does not need to be handled. You can run the **display ring check config** command to check whether the ring detection function is enabled.
- In the aggregation management scenario, the Ethernet port configurations of a slave OLT for the master/slave connection are subject to the default configurations, the configurations made on the master OLT for a slave OLT do not take effect, and the displayed Ethernet port configurations may be different from the actual configurations that take effect.

Step 1 Configure the forwarding VLAN and connection ports between the master and slave OLTs.

1. Run the **vlan** command to add a VLAN.

   NOTE

   The VLAN attribute cannot be QinQ or stacking. It is recommended that you use the smart VLAN.

2. Run the **port vlan** command on the master OLT to add ports connected to slave OLTs to the VLAN.

NOTICE

When configuring or deleting reserved VLANs for aggregation management on the master OLT by running a command, ensure that the number of reserved VLANs for aggregation management is greater than or equal to the total number of non-reserved VLANs after the command is executed. You can run the **display vlan all** command to query the total number of non-reserved VLANs.

 Copyright © Huawei Technologies Co., Ltd.

NOTE

Aggregation management can be configured in customized configuration mode (steps 2 to 4) or one-click configuration mode (step 5). The two modes cannot be configured at the same time.

If you want to customize the network plan, you are advised to perform steps 2 to 4. If you want to simplify the configuration, go to step 5.

Step 2 Configure the service attributes of the forwarding VLANs of the master and slave OLTs.

1. Run the **vlan service-profile** command to create a VLAN service profile.

   NOTE

   It is used to distinguish the service attributes of underlay VLANs used for aggregation management from those of other VLANs, such as the VLAN forwarding policy and forwarding mode, thereby preventing the global configuration from interfering with aggregation management.

2. Run the **dhcp mode** command to specify the DHCP forwarding mode.
3. Run the **dhcp option82** command to disable the DHCP Option 82 function.

   NOTE

   The DHCP server cannot coexist with Option 82.

4. Run the **vlan bind service-profile** command to bind the VLAN service profile to a VLAN.

Step 3 Configure a DHCP server.

1. Run the **interface vlanif** command to create a VLANIF interface and enter the VLANIF mode.
2. Run the **ip address** command to configure the IP address of the DHCP server.
3. Run the **dhcp select interface** command to enable the DHCP server.
4. Run the **dhcp server ip-range** command to configure the address pool range on the DHCP server.

NOTICE

The IP address of the DHCP server cannot be in the same network segment as the aggregation management IP address.

Step 4 Configure reserved resources for aggregation management.

1. Run the **co resource bridge-domain** command to configure the BD resources for aggregation management on the master OLT.
2. Run the **co resource vni** command to configure VXLAN network identifier (VNI) resources for aggregation management on the master OLT.
3. Run the **co resource management-ip** command to configure the IP address for aggregation management on the master OLT.

Step 5 Configure aggregation management in batches in one-click mode. (If steps 2 to 4 are not performed, perform this step.)

 Copyright © Huawei Technologies Co., Ltd.

1. Run the **co enable vlan** *vlan-id* **interface-ip** *ip-address* **management-ip** *ip-address* command to configure aggregation management in batches in one-click mode. After this command is executed, a VLAN service profile for aggregation management is generated and bound to a VLAN. A Layer 3 interface is generated and configured with an IP address. The DHCP server is enabled, and the reserved resources and management IP address are configured for aggregation management.

   NOTE

   If a configuration conflict occurs in the system, the command fails to be executed. Delete the conflicting configuration or perform steps 2 to 4 to use the customized configuration mode.

   The automatically allocated aggregation management configurations are as follows:

   1. VLAN service profile ID: 256; name: srvprof-co-management
   2. Reserved BD resources of the master OLT (range: 8192–12255)
   3. Reserved VNI resources of the master OLT (range: 32768–36831)

Step 6 (Optional) Configure aggregation protection for master and slave OLTs.

1. Run the **dhcp server option** command to configure user-defined options for the DHCP server address pool.

   NOTE

   If the port connecting master and slave OLTs is expected to work in aggregation protection mode, run the **dhcp server option** command on the master OLT to specify the port aggregation mode of a slave OLT.

   The port aggregation mode of a slave OLT is as follows:

   1. **dhcp server option 226 ascii 01** indicates that ports 0 and 1 on the board connecting a slave OLT to the master OLT are configured as an aggregation group.
   2. **dhcp server option 226 ascii 0123** indicates that ports 0, 1, 2, and 3 on the board connecting a slave OLT to the master OLT are configured as an aggregation group.

   In V100R023C0x and later versions, the **co resource link-aggregation port** { **half** | **full** } [**workmode lacp-static**] command can be used to replace the **dhcp server option** command. The new command supports the LACP aggregation group for slave OLTs. If both of commands are run, the configuration of the **co resource link-aggregation port** command takes effect.

   1. **co resource link-aggregation port half** is equivalent to **dhcp server option 226 ascii 01**.
   2. **co resource link-aggregation port full** is equivalent to **dhcp server option 226 ascii 0123**.

2. Run the **port vlan** command to add the ports to be aggregated on the master and slave OLTs to the forwarding VLAN.

   NOTE

   Skip this configuration if it has been performed in step 1.

3. Run the **link-aggregation** command to create an Ethernet link aggregation group (LAG) and specify the master port of the LAG.
4. Run the **link-aggregation add-member** command to add an Ethernet port to the existing LAG.

   NOTE

   After the LAG is configured, you can run the **display co current-configuration** command to query the related configurations.

 Copyright © Huawei Technologies Co., Ltd.

Step 7 Check the online status of slave OLTs.

1. Run the **display frame info** command to query the online status of slave OLTs.

NOTE

After a slave OLT goes online, the local serial port and Meth Ethernet port are disabled.

----End

21.7.3 Configuration Example

21.7.3.1 Customized Configuration Example

Prerequisites

- The master and slave OLTs have been installed and powered on.
- The software versions of the master and slave OLTs are the same and support the aggregation management feature.
- A slave OLT must support certificate pre-configuration. The master OLT must support certificate pre-configuration or manual digital certificate loading. For details about how to query and load a digital certificate, see **1.7.1 Process of Loading a Digital Certificate**.
- A slave OLT must be started with an empty database and cannot be logged in after the startup. Otherwise, the slave OLT will be restored to the independent mode.
- It is recommended that aggregation management be used with NCE. Currently, third-party OSSs and U2000 do not fully support aggregation management.
- Do not use the centralized management and slicing features at the same time on the master OLT. Otherwise, services on the slave OLTs will be abnormal.
- When configuring aggregation management on the master OLT, ensure that DCN is disabled globally. You can run the **display dcn global** command to check whether DCN is enabled globally. If DCN is enabled globally, run the **undo dcn** command to disable it. If STP has been enabled globally, disable STP on the port connected to a slave OLT to prevent intermittent link disconnection caused by STP calculation. You can run the **stp port disable** command to disable STP.

  NOTE

  The master OLT can be interconnected with a slave OLT only through a GE/10GE port. An xPON port cannot be used for interconnection. To ensure bandwidth, it is recommended that at least 10 Gbit/s interface boards be used on the network. If the bandwidth is insufficient, the OLT may be unreachable by the NMS. This is not a mandatory requirement. Customers need to make configurations based on actual requirements.

 Copyright © Huawei Technologies Co., Ltd.

Data Planning

Configuration Item	Configuration on the Master OLT
Aggregation management port	Master OLT: 0/9/0, 0/10/0, 0/9/1, and 0/10/1 Slave OLT 1: 1/2/0, 1/2/1 Slave OLT 2: 2/2/0, 2/2/1
VXLAN tunnel resources of the master OLT	BD resources: 64–95 VNI resources: 64–95
Aggregation management IP address of the master OLT	10.0.1.1
Aggregation management VLAN	VLAN ID: 2
Service profile of the aggregation management VLAN	ID: 1 Name: co+x
IP address of the DHCP server	10.0.0.1/24
IP address pool range of the DHCP server	10.0.0.1 to 10.0.0.100

Configuration Procedure

1. Create a VLAN with the ID of 2.
   ```
   huawei(config)#vlan 2
   ```
2. Create a VLAN service profile named co+x with the ID of 1.
   ```
   huawei(config)#vlan service-profile profile-id 1 profile-name "co+x"
   ```
3. Set the DHCP forwarding mode to Layer 2 forwarding.
   ```
   huawei(config-vlan-srvprof-1)#dhcp mode layer-2
   ```
4. Disable the DHCP Option 82 function.
   ```
   huawei (config-vlan-srvprof-1)#dhcp option82 disable
   huawei (config-vlan-srvprof-1)#commit
   huawei (config-vlan-srvprof-1)#quit
   ```
5. Bind VLAN service profile 1 to VLAN 2.
   ```
   huawei(config)#vlan bind service-profile 2 profile-id 1
   ```
6. Create VLANIF interface 2 and enter the VLANIF mode.
   ```
   huawei(config)#interface vlanif 2
   ```
7. Set the IP address of the DHCP server to 10.0.0.1.
   ```
   huawei(config-if-vlanif2)#ip address 10.0.0.1 24
   ```
8. Enable the DHCP server function.
   ```
   huawei(config-if-vlanif2)#dhcp select interface
   ```
9. Set the IP address pool range of the DHCP server to 10.0.0.1–10.0.0.100.
   ```
   huawei(config-if-vlanif2)#dhcp server ip-range 10.0.0.1 10.0.0.100
   ```
10. Set Option 226 of the DHCP server to 01.
    ```
    huawei(config-if-vlanif2)#dhcp server option 226 ascii 01
    huawei(config-if-vlanif2)#quit
    ```
11. Create aggregation groups for master OLT ports 0/9/0 and 0/9/1 that are connected to slave OLTs, and set the packet distribution mode to distributing packets according to the source MAC address and destination MAC address.

 Copyright © Huawei Technologies Co., Ltd.

```
huawei(config)#link-aggregation 0/9 0 egress-ingress
huawei(config)#link-aggregation 0/9 1 egress-ingress
```

12. Add port 0/10/0 of the master OLT to the aggregation group of port 0/9/0, and add port 0/10/1 to the aggregation group of port 0/9/1.

```
huawei(config)#link-aggregation add-member 0/9/0 0/10 0
huawei(config)#link-aggregation add-member 0/9/1 0/10 1
```

13. Configure ports 0/9/0 and 0/9/1 of the master OLT as upstream ports of the aggregation management VLAN.

```
huawei(config)#port vlan 2 0/9 0
huawei(config)#port vlan 2 0/9 1
```

14. Set the ID range of reserved BD resources to 64–95 for the master OLT.

```
huawei(config)#co resource bridge-domain 64 session-num 1
```

15. Set the ID range of reserved VNI resources to 64–95 for the master OLT.

```
huawei(config)#co resource vni 64 session-num 1
```

16. Set the aggregation management IP address to 10.0.1.1 for the master OLT.

```
huawei(config)#co resource management-ip 10.0.1.1
huawei(config)#quit
```

17. Check the online status of a slave OLT after aggregation management is configured on the master OLT. If a slave OLT is in the **Normal** state, it is successfully connected with the master OLT.

```
huawei(config)#display frame info
{<cr>|bydesc<K>|frameid<U><0,15>||<K> }:

 Command:
        display frame info
 ------------------------------------------------------------
 FrameID      FrameType          FrameState
Role
 ------------------------------------------------------------
 0            H901BPIB           Normal      Active
 1            MA5801-GP08-H2     Normal      Extended
 2            MA5801-GP08-H2     Normal      Extended
 ------------------------------------------------------------
 Local Config FrameID: 0
 ------------------------------------------------------------
 Total: 3
```

Configuration Script

```
vlan 2
vlan service-profile profile-id 1 profile-name "co+x"
dhcp mode layer-2
dhcp option82 disable
commit
quit
vlan bind service-profile 2 profile-id 1
interface vlanif 2
ip address 10.0.0.1 24
dhcp select interface
dhcp server ip-range 10.0.0.1 10.0.0.100
dhcp server option 226 ascii 01
quit
link-aggregation 0/9 0 egress-ingress
link-aggregation 0/9 1 egress-ingress
link-aggregation add-member 0/9/0 0/10 0
link-aggregation add-member 0/9/1 0/10 1
port vlan 2 0/9 0
port vlan 2 0/9 1
co resource bridge-domain 64 session-num 1
co resource vni 64 session-num 1
co resource management-ip 10.0.1.1
quit
display frame info
```

 Copyright © Huawei Technologies Co., Ltd.

21.7.3.2 Batch Configuration Example

Prerequisites

- The master and slave OLTs have been installed and powered on.
- The software versions of the master and slave OLTs are the same and support the aggregation management feature.
- A slave OLT must support certificate pre-configuration. The master OLT must support certificate pre-configuration or manual digital certificate loading. For details about how to query and load a digital certificate, see **1.7.1 Process of Loading a Digital Certificate**.
- A slave OLT must be started with an empty database and cannot be logged in after the startup. Otherwise, the slave OLT will be restored to the independent mode.
- It is recommended that aggregation management be used with NCE. Currently, third-party OSSs and U2000 do not fully support aggregation management.
- Do not use the centralized management and slicing features at the same time on the master OLT. Otherwise, services on the slave OLTs will be abnormal.
- When configuring aggregation management on the master OLT, ensure that DCN is disabled globally. You can run the **display dcn global** command to check whether DCN is enabled globally. If DCN is enabled globally, run the **undo dcn** command to disable it. If STP has been enabled globally, disable STP on the port connected to a slave OLT to prevent intermittent link disconnection caused by STP calculation. You can run the **stp port disable** command to disable STP.

  NOTE

  The master OLT can be interconnected with a slave OLT only through a GE/10GE port. An xPON port cannot be used for interconnection. To ensure bandwidth, it is recommended that at least 10 Gbit/s interface boards be used on the network. If the bandwidth is insufficient, the OLT may be unreachable by the NMS. This is not a mandatory requirement. Customers need to make configurations based on actual requirements.

Data Planning

Configuration Item	Configuration on the Master OLT
Aggregation management port	Master OLT: 0/9/0, 0/10/0, 0/9/1, and 0/10/1 Slave OLT 1: 1/2/0, 1/2/1 Slave OLT 2: 2/2/0, 2/2/1
Aggregation management IP address of the master OLT	10.0.1.1
Aggregation management VLAN	VLAN ID: 2
IP address of the DHCP server	10.0.0.1/24

 Copyright © Huawei Technologies Co., Ltd.

Configuration Procedure

1. Create a VLAN with the ID of 2.
   ```
   huawei(config)#vlan 2
   ```
2. Configure ports 0/9/0 and 0/9/1 of the master OLT as upstream ports of the aggregation management VLAN.
   ```
   huawei(config)#port vlan 2 0/9 0
   huawei(config)#port vlan 2 0/9 1
   ```
3. Configure a VLAN service profile for aggregation management and bind it to VLAN 2. Create a Layer 3 interface, set the IP address of the DHCP server to 10.0.0.1, enable the DHCP server, and set the aggregation management IP address of the master OLT to 10.0.0.0.
   ```
   huawei(config)#co enable vlan 2 interface ip 10.0.0.1 management ip 10.0.1.1
   ```
4. Create VLANIF interface 2 and enter the VLANIF mode.
   ```
   huawei(config)#interface vlanif 2
   ```
5. Set Option 226 of the DHCP server to 01.
   ```
   huawei(config-if-vlanif2)#dhcp server option 226 ascii 01
   huawei(config-if-vlanif2)#quit
   ```
6. Create aggregation groups for master OLT ports 0/9/0 and 0/9/1 that are connected to slave OLTs, and set the packet distribution mode to distributing packets according to the source MAC address and destination MAC address.
   ```
   huawei(config)#link-aggregation 0/9 0 egress-ingress
   huawei(config)#link-aggregation 0/9 1 egress-ingress
   ```
7. Add port 0/10/0 of the master OLT to the aggregation group of port 0/9/0, and add port 0/10/1 to the aggregation group of port 0/9/1.
   ```
   huawei(config)#link-aggregation add-member 0/9/0 0/10 0
   huawei(config)#link-aggregation add-member 0/9/1 0/10 1
   ```
8. Check the online status of a slave OLT after aggregation management is configured on the master OLT. If a slave OLT is in the **Normal** state, it is successfully connected with the master OLT.
   ```
   huawei(config)#display frame info
   {<cr>|bydesc<K>|frameid<U><0,15>||<K> }:

    Command:
          display frame info
    ------------------------------------------------------------
    FrameID      FrameType          FrameState
   Role
    ------------------------------------------------------------
    0            H901BPIB           Normal      Active
    1            MA5801-GP08-H2     Normal      Extended
    2            MA5801-GP08-H2     Normal      Extended
    ------------------------------------------------------------
    Local Config FrameID: 0
    ------------------------------------------------------------
    Total: 3
   ```

Configuration Script

```
vlan 2
port vlan 2 0/9 0
port vlan 2 0/9 1
co enable vlan 2 interface ip 10.0.0.1 management ip 10.0.1.1
interface vlanif 2
dhcp server option 226 ascii 01
quit
link-aggregation 0/9 0 egress-ingress
link-aggregation 0/9 1 egress-ingress
link-aggregation add-member 0/9/0 0/10 0
```

 Copyright © Huawei Technologies Co., Ltd.

```
link-aggregation add-member 0/9/1 0/10 1
display frame info
```