



3Com[®] Switch 4800G Family

Configuration Guide

Switch 4800G 24-Port
Switch 4800G PWR 24-Port
Switch 4800G 48-Port
Switch 4800G PWR 48-Port
Switch 4800G 24-Port SFP

www.3Com.com
Part Number: 10015265 Rev. AB
Published: March 2008

3Com Corporation
350 Campus Drive
Marlborough, MA
USA 01752-3064

Copyright © 2006-2008, 3Com Corporation. All rights reserved. No part of this documentation may be reproduced in any form or by any means or used to make any derivative work (such as translation, transformation, or adaptation) without written permission from 3Com Corporation.

3Com Corporation reserves the right to revise this documentation and to make changes in content from time to time without obligation on the part of 3Com Corporation to provide notification of such revision or change.

3Com Corporation provides this documentation without warranty, term, or condition of any kind, either implied or expressed, including, but not limited to, the implied warranties, terms or conditions of merchantability, satisfactory quality, and fitness for a particular purpose. 3Com may make improvements or changes in the product(s) and/or the program(s) described in this documentation at any time.

If there is any software on removable media described in this documentation, it is furnished under a license agreement included with the product as a separate document, in the hard copy documentation, or on the removable media in a directory file named LICENSE.TXT or !LICENSE.TXT. If you are unable to locate a copy, please contact 3Com and a copy will be provided to you.

UNITED STATES GOVERNMENT LEGEND

If you are a United States government agency, then this documentation and the software described herein are provided to you subject to the following:

All technical data and computer software are commercial in nature and developed solely at private expense. Software is delivered as "Commercial Computer Software" as defined in DFARS 252.227-7014 (June 1995) or as a "commercial item" as defined in FAR 2.101(a) and as such is provided with only such rights as are provided in 3Com's standard commercial license for the Software. Technical data is provided with limited rights only as provided in DFAR 252.227-7015 (Nov 1995) or FAR 52.227-14 (June 1987), whichever is applicable. You agree not to remove or deface any portion of any legend provided on any licensed program or documentation contained in, or delivered to you in conjunction with, this User Guide.

Unless otherwise indicated, 3Com registered trademarks are registered in the United States and may or may not be registered in other countries.

3Com and the 3Com logo are registered trademarks of 3Com Corporation.

Cisco is a registered trademark of Cisco Systems, Inc.

Funk RADIUS is a registered trademark of Funk Software, Inc.

Aegis is a registered trademark of Aegis Group PLC.

Intel and Pentium are registered trademarks of Intel Corporation. Microsoft, MS-DOS, Windows, and Windows NT are registered trademarks of Microsoft Corporation. Novell and NetWare are registered trademarks of Novell, Inc. UNIX is a registered trademark in the United States and other countries, licensed exclusively through X/Open Company, Ltd.

IEEE and 802 are registered trademarks of the Institute of Electrical and Electronics Engineers, Inc.

All other company and product names may be trademarks of the respective companies with which they are associated.

ENVIRONMENTAL STATEMENT

It is the policy of 3Com Corporation to be environmentally-friendly in all operations. To uphold our policy, we are committed to:

Establishing environmental performance standards that comply with national legislation and regulations.

Conserving energy, materials and natural resources in all operations.

Reducing the waste generated by all operations. Ensuring that all waste conforms to recognized environmental standards. Maximizing the recyclable and reusable content of all products.

Ensuring that all products can be recycled, reused and disposed of safely.

Ensuring that all products are labelled according to recognized environmental standards.

Improving our environmental record on a continual basis.

End of Life Statement

3Com processes allow for the recovery, reclamation and safe disposal of all end-of-life electronic components.

Regulated Materials Statement

3Com products do not contain any hazardous or ozone-depleting material.

CONTENTS

ABOUT THIS GUIDE

- Conventions 21
- Related Documentation 21

PRODUCT OVERVIEW

- Preface 23
- Product Models 23

NETWORKING APPLICATIONS

- Serving as a Convergence Layer Device 24
- Serving as a Access Layer Device 24

1 LOGGING IN TO AN ETHERNET SWITCH

- Logging In to an Ethernet Switch 27
- Introduction to the User Interface 27

2 LOGGING IN THROUGH THE CONSOLE PORT

- Introduction 31
- Setting Up the Connection to the Console Port 31
- Console Port Login Configuration 33
- Console Port Login Configuration with Authentication Mode Being None 35
- Console Port Login Configuration with Authentication Mode Being Password 38
- Console Port Login Configuration with Authentication Mode Being Scheme 41

3 LOGGING IN THROUGH TELNET

- Introduction 47
- Telnet Configuration with Authentication Mode Being None 49
- Telnet Configuration with Authentication Mode Being Password 52
- Telnet Configuration with Authentication Mode Being Scheme 54
- Telnet Connection Establishment 59

4 LOGGING IN USING MODEM

- Introduction 63
- Configuration on the Administrator Side 63
- Configuration on the Switch Side 63
- Modem Connection Establishment 64

5 LOGGING IN THROUGH WEB-BASED NETWORK MANAGEMENT SYSTEM

- Introduction 67
- HTTP Connection Establishment 67
- Web Server Shutdown/Startup 68
- Displaying Web Users 69

6 LOGGING IN THROUGH NMS

- Introduction 71
- Connection Establishment Using NMS 71

7 CONFIGURING SOURCE IP ADDRESS FOR TELNET SERVICE PACKETS

- Overview 73
- Configuring Source IP Address for Telnet Service Packets 73
- Displaying the source IP address/Interface Specified for Telnet Packets 74

8 CONTROLLING LOGIN USERS

- Introduction 75
- Controlling Telnet Users 75
- Controlling Network Management Users by Source IP Addresses 78
- Controlling Web Users by Source IP Address 79

9 VLAN CONFIGURATION

- Introduction to VLAN 83
- Configuring Basic VLAN Attributes 86
- Basic VLAN Interface Configuration 86
- Port-Based VLAN Configuration 87
- MAC Address-Based VLAN Configuration 91
- Protocol-Based VLAN Configuration 92
- Configuring IP-Subnet-Based VLAN 94
- Displaying and Maintaining VLAN 95
- VLAN Configuration Example 95

10 VOICE VLAN CONFIGURATION

- Introduction to Voice VLAN 99
- Configuring Voice VLAN 101
- Displaying and Maintaining Voice VLAN 103
- Voice VLAN Configuration Examples 103

11 GVRP CONFIGURATION

- Introduction to GVRP 109
- GVRP Configuration Task List 112
- Configuring GVRP 112
- Displaying and Maintaining GVRP 114
- GVRP Configuration Examples 114

12 IP ADDRESSING CONFIGURATION

- IP Addressing Overview 121
- Configuring IP Addresses 123
- Displaying and Maintaining IP Addressing 126

13 IP PERFORMANCE CONFIGURATION

- IP Performance Overview 127
- Enabling Reception and Forwarding of Directed Broadcasts to a Directly Connected Network 127
- Configuring TCP Attributes 129
- Configuring ICMP to Send Error Packets 130
- Displaying and Maintaining IP Performance 132

14 QINQ CONFIGURATION

- Introduction to QinQ 133
- Configuring Basic QinQ 135
- Configuring Selective QinQ 136
- Configuring the TPID Value to Be Carried in VLAN Tags 137
- QinQ Configuration Example 137

15 BPDU TUNNELING CONFIGURATION

- Introduction to BPDU Tunneling 141
- Configuring BPDU Isolation 142
- Configuring BPDU Transparent Transmission 143
- Configuring Destination Multicast MAC Address for BPDU Tunnel Frames 144
- BPDU Tunneling Configuration Example 144

16 PORT CORRELATION CONFIGURATION

- Ethernet Port Configuration 147
- Maintaining and Displaying an Ethernet Port 156

17 PORT ISOLATION CONFIGURATION

- Introduction to Port Isolation 157
- Configuring an Isolation Group 157
- Displaying Isolation Groups 158
- Port Isolation Configuration Example 158

18 LINK AGGREGATION OVERVIEW

- Link Aggregation 161
- Approaches to Link Aggregation 162
- Load Sharing in a Link Aggregation Group 165
- Service Loop Group 165
- Aggregation Port Group 166

19 LINK AGGREGATION CONFIGURATION

- Configuring Link Aggregation 167
- Displaying and Maintaining Link Aggregation 169
- Link Aggregation Configuration Example 170

20 MAC ADDRESS TABLE MANAGEMENT CONFIGURATION

- Introduction to MAC Address Table 173
- Configuring MAC Address Table Management 174
- Displaying and Maintaining MAC Address Table Management 176
- MAC Address Table Management Configuration Example 176

21 IP SOURCE GUARD CONFIGURATION

- IP Source Guard Overview 177
- Configuring a Static Binding Entry 177
- Configuring Dynamic Binding Function 178
- Displaying IP Source Guard 178
- IP Source Guard Configuration Examples 178
- Troubleshooting 182

22 DLDP CONFIGURATION

- Overview 183
- DLDP Configuration Task List 190
- Displaying and Maintaining DLDP 193
- DLDP Configuration Example 193
- Troubleshooting 195

23 MSTP CONFIGURATION

- MSTP Overview 197
- Configuration Task List 212
- Configuring the Root Bridge 213
- Configuring Leaf Nodes 224
- Performing mCheck 228
- Configuring Digest Snooping 229
- Configuring No Agreement Check 230
- Configuring Protection Functions 233
- Displaying and Maintaining MSTP 235
- MSTP Configuration Example 236

24 IP ROUTING OVERVIEW

- IP Routing and Routing Table 241
- Routing Protocol Overview 243
- Displaying and Maintaining a Routing Table 246

25	GR OVERVIEW	
	Introduction to Graceful Restart	247
	Basic Concepts in Graceful Restart	247
	Graceful Restart Communication Procedure	248
	Graceful Restart Mechanism for Several Commonly Used Protocols	250
26	STATIC ROUTING CONFIGURATION	
	Introduction	251
	Configuring a Static Route	252
	Detecting Reachability of the Static Route's Nexthop	253
	Displaying and Maintaining Static Routes	254
	Configuration Example	254
27	RIP CONFIGURATION	
	RIP Overview	257
	Configuring RIP Basic Functions	261
	Configuring RIP Route Control	263
	Configuring RIP Network Optimization	266
	Displaying and Maintaining RIP	269
	RIP Configuration Examples	269
	Troubleshooting RIP	271
28	OSPF CONFIGURATION	
	Introduction to OSPF	273
	OSPF Configuration Task List	292
	Configuring OSPF Basic Functions	293
	Configuring OSPF Area Parameters	294
	Configuring OSPF Network Types	295
	Configuring OSPF Route Control	297
	Configuring OSPF Network Optimization	300
	Configuring OSPF Graceful Restart	306
	Displaying and Maintaining OSPF	309
	OSPF Configuration Examples	309
	Troubleshooting OSPF Configuration	323
29	IS-IS CONFIGURATION	
	IS-IS Overview	325
	IS-IS Configuration Task List	340
	Configuring IS-IS Basic Functions	341
	Configuring IS-IS Routing Information Control	342
	Tuning and Optimizing IS-IS Network	346
	Configuring IS-IS GR	352
	Displaying and Maintaining IS-IS	353
	IS-IS Configuration Example	354

30 BGP CONFIGURATION

- BGP Overview 365
- BGP Configuration Task List 380
- Configuring BGP Basic Functions 381
- Controlling Route Distribution and Reception 383
- Configuring BGP Route Attributes 386
- Tuning and Optimizing BGP Networks 388
- Configuring a Large Scale BGP Network 390
- Configuring BGP GR 392
- Displaying and Maintaining BGP 394
- BGP Configuration Examples 395
- Troubleshooting BGP 413

31 ROUTING POLICY CONFIGURATION

- Introduction to Routing Policy 415
- Routing Policy Configuration Task List 417
- Defining Filtering Lists 417
- Configuring a Routing Policy 419
- Displaying and Maintaining the Routing Policy 422
- Routing Policy Configuration Example 422
- Troubleshooting Routing Policy Configuration 425

32 IPV6 STATIC ROUTING CONFIGURATION

- Introduction to IPv6 Static Routing 427
- Configuring an IPv6 Static Route 427
- Displaying and Maintaining IPv6 Static Routes 428
- IPv6 Static Routing Configuration Example 428

33 IPV6 RIPNG CONFIGURATION

- Introduction to RIPng 431
- Configuring RIPng Basic Functions 433
- Configuring RIPng Route Control 434
- Tuning and Optimizing the RIPng Network 436
- Displaying and Maintaining RIPng 438
- RIPng Configuration Example 438

34 IPV6 OSPFV3 CONFIGURATION

- Introduction to OSPFv3 443
- IPv6 OSPFv3 Configuration Task List 445
- Configuring OSPFv3 Basic Functions 446
- Configuring OSPFv3 Area Parameters 446
- Configuring OSPFv3 Routing Information Management 447
- Tuning and Optimizing an OSPFv3 Network 450
- Displaying and Maintaining OSPFv3 452
- OSPFv3 Configuration Examples 453

Troubleshooting OSPFv3 Configuration 459

35 IPv6 IS-IS CONFIGURATION

Introduction to IPv6 IS-IS 461
Configuring IPv6 IS-IS Basic Functions 461
Configuring IPv6 IS-IS Routing Information Control 462
Displaying and Maintaining IPv6 IS-IS 463
IPv6 IS-IS Configuration Example 464

36 IPv6 BGP CONFIGURATION

IPv6 BGP Overview 467
Configuration Task List 468
Configuring IPv6 BGP Basic Functions 469
Controlling Route Distribution and Reception 471
Configuring IPv6 BGP Route Attributes 474
Tuning and Optimizing IPv6 BGP Networks 476
Configuring a Large Scale IPv6 BGP Network 478
Displaying and Maintaining IPv6 BGP Configuration 482
IPv6 BGP Configuration Examples 483
Troubleshooting IPv6 BGP Configuration 486

37 ROUTING POLICY CONFIGURATION

Introduction to Routing Policy 489
Defining Filtering Lists 490
Configuring a Routing Policy 492
Displaying and Maintaining the Routing Policy 495
Routing Policy Configuration Example 495
Troubleshooting Routing Policy Configuration 497

38 IPv6 BASICS CONFIGURATION

IPv6 Overview 499
IPv6 Basics Configuration Task List 508
Configuring Basic IPv6 Functions 508
Configuring IPv6 NDP 510
Configuring PMTU Discovery 513
Configuring IPv6 TCP Properties 514
Configuring ICMPv6 Packet Sending 514
Configuring IPv6 DNS 515
Displaying and Maintaining IPv6 Basics Configuration 516
IPv6 Configuration Example 517
Troubleshooting IPv6 Basics Configuration 520

39 DUAL STACK CONFIGURATION

Dual Stack Overview 521
Configuring Dual Stack 521

40 TUNNELING CONFIGURATION

- Introduction to Tunneling 523
- Tunneling Configuration Task List 526
- Configuring IPv6 Manual Tunnel 526
- Configuring 6to4 Tunnel 530
- Configuring ISATAP Tunnel 535
- Displaying and Maintaining Tunneling Configuration 538
- Troubleshooting Tunneling Configuration 538

41 MULTICAST OVERVIEW

- Introduction to Multicast 541
- Multicast Models 544
- Multicast Architecture 545
- Multicast Packet Forwarding Mechanism 550

42 IGMP SNOOPING CONFIGURATION

- IGMP Snooping Overview 553
- IGMP Snooping Configuration Task List 558
- Configuring Basic Functions of IGMP Snooping 559
- Configuring IGMP Snooping Port Functions 560
- Configuring IGMP Snooping Querier 563
- Configuring an IGMP Snooping Policy 565
- Displaying and Maintaining IGMP Snooping 569
- IGMP Snooping Configuration Examples 570
- Troubleshooting IGMP Snooping Configuration 577

43 MLD SNOOPING CONFIGURATION

- MLD Snooping Overview 579
- MLD Snooping Configuration Task List 583
- Configuring Basic Functions of MLD Snooping 584
- Configuring MLD Snooping Port Functions 585
- Configuring MLD Snooping Querier 589
- Configuring an MLD Snooping Policy 591
- Displaying and Maintaining MLD Snooping 595
- MLD Snooping Configuration Examples 596
- Troubleshooting MLD Snooping 602

44 MULTICAST VLAN CONFIGURATION

- Introduction to Multicast VLAN 605
- Configuring Multicast VLAN 605
- Displaying and Maintaining Multicast VLAN 606
- Multicast VLAN Configuration Example 606

45 IPV6 MULTICAST VLAN CONFIGURATION

- Introduction to IPv6 Multicast VLAN 609

Configuring IPv6 Multicast VLAN 609
Displaying and Maintaining IPv6 Multicast VLAN 610
IPv6 Multicast VLAN Configuration Examples 610

46 IGMP CONFIGURATION

IGMP Overview 613
IGMP Configuration Task List 617
Configuring Basic Functions of IGMP 618
Adjusting IGMP Performance 620
Displaying and Maintaining IGMP 623
IGMP Configuration Example 624
Troubleshooting IGMP 626

47 PIM CONFIGURATION

PIM Overview 629
Configuring PIM-DM 641
Configuring PIM-SM 643
Configuring PIM-SSM 652
Configuring PIM Common Information 653
Displaying and Maintaining PIM 658
PIM Configuration Examples 659
Troubleshooting PIM Configuration 669

48 MSDP CONFIGURATION

MSDP Overview 673
MSDP Configuration Task List 679
Configuring Basic Functions of MSDP 679
Configuring an MSDP Peer Connection 680
Configuring SA Messages Related Parameters 682
Displaying and Maintaining MSDP 685
MSDP Configuration Examples 685
Troubleshooting MSDP 697

49 MULTICAST ROUTING AND FORWARDING CONFIGURATION

Multicast Routing and Forwarding Overview 701
Configuration Task List 705
Configuring Multicast Routing and Forwarding 706
Displaying and Maintaining Multicast Routing and Forwarding 709
Configuration Examples 709
Troubleshooting Multicast Routing and Forwarding 713

50 802.1X CONFIGURATION

802.1x Overview 715
Configuring 802.1x 726
Configuring a Guest VLAN 728

Displaying and Maintaining 802.1x 729
802.1x Configuration Example 729
Guest VLAN Configuration Example 732
ACL Assignment Configuration Example 735

51 HABP CONFIGURATION

Introduction to HABP 737
Configuring HABP 737
Displaying and Maintaining HABP 738

52 MAC AUTHENTICATION CONFIGURATION

MAC Authentication Overview 739
Related Concepts 740
Configuring MAC Authentication 741
Displaying and Maintaining MAC Authentication 742
MAC Authentication Configuration Examples 742

53 AAA/RADIUS/HWTACACS CONFIGURATION

AAA/RADIUS/HWTACACS Overview 747
AAA/RADIUS/HWTACACS Configuration Task List 756
Configuring AAA 758
Configuring RADIUS 765
Configuring HWTACACS 771
Displaying and Maintaining AAA/RADIUS/HWTACACS 775
AAA/RADIUS/HWTACACS Configuration Examples 776
Troubleshooting AAA/RADIUS/HWTACACS 779

54 ARP CONFIGURATION

ARP Overview 781
Configuring ARP 783
Configuring Gratuitous ARP 785
Displaying and Maintaining ARP 786

55 PROXY ARP CONFIGURATION

Proxy ARP Overview 787
Enabling Proxy ARP 787
Displaying and Maintaining Proxy ARP 787
Proxy ARP Configuration Examples 788

56 DHCP OVERVIEW

Introduction to DHCP 791
DHCP Address Allocation 792
DHCP Message Format 793
DHCP Options 794

57 DHCP SERVER CONFIGURATION

- Introduction to DHCP Server 797
- DHCP Server Configuration Task List 799
- Enabling DHCP 799
- Enabling the DHCP Server on an Interface 799
- Configuring an Address Pool for the DHCP Server 800
- Configuring the DHCP Server Security Functions 806
- Configuring the Handling Mode for Option 82 808
- Displaying and Maintaining the DHCP Server 808
- DHCP Server Configuration Examples 809
- Troubleshooting DHCP Server Configuration 811

58 DHCP RELAY AGENT CONFIGURATION

- Introduction to DHCP Relay Agent 813
- Configuration Task List 815
- Configuring the DHCP Relay Agent 815
- Displaying and Maintaining DHCP Relay Agent Configuration 819
- DHCP Relay Agent Configuration Example 820
- Troubleshooting DHCP Relay Agent Configuration 821

59 DHCP CLIENT CONFIGURATION

- Introduction to DHCP Client 823
- Enabling the DHCP Client on an Interface 823
- Displaying and Maintaining the DHCP Client 824
- DHCP Client Configuration Example 824

60 DHCP SNOOPING CONFIGURATION

- DHCP Snooping Overview 825
- Configuring DHCP Snooping Basic Functions 828
- Configuring DHCP Snooping to Support Option 82 828
- Displaying and Maintaining DHCP Snooping 829
- DHCP Snooping Configuration Example 829

61 BOOTP CLIENT CONFIGURATION

- Introduction to BOOTP Client 831
- Configuring an Interface to Dynamically Obtain an IP Address Through BOOTP 832
- Displaying and Maintaining BOOTP Client Configuration 832
- BOOTP Client Configuration Example 832

62 ACL OVERVIEW

- Introduction to ACL 835
- Introduction to IPv4 ACL 836

Introduction to IPv6 ACL 838

63 IPv4 ACL CONFIGURATION

Creating a Time Range 841
Configuring a Basic IPv4 ACL 842
Configuring an Advanced IPv4 ACL 844
Configuring an Ethernet Frame Header ACL 845
Copying an IPv4 ACL 846
Displaying and Maintaining IPv4 ACLs 847
IPv4 ACL Configuration Example 847

64 IPv6 ACL CONFIGURATION

Creating a Time Range 851
Configuring a Basic IPv6 ACL 851
Configuring an Advanced IPv6 ACL 852
Copying an IPv6 ACL 854
Displaying and Maintaining IPv6 ACLs 854
IPv6 ACL Configuration Example 854

65 QoS OVERVIEW

Introduction 857
Traditional Packet Forwarding Service 857
New Requirements Brought forth by New Services 857
Occurrence and Influence of Congestion and the Countermeasures 858
Major Traffic Management Techniques 859

66 TRAFFIC CLASSIFICATION, TP, AND LR CONFIGURATION

Traffic Classification Overview 861
TP and LR Overview 864
Traffic Evaluation and Token Bucket 864
LR Configuration 866
Displaying and Maintaining LR 867

67 QoS POLICY CONFIGURATION

Overview 869
Configuring QoS Policy 870
Introduction to QoS Policies 870
Configuring a QoS Policy 870
Displaying and Maintaining QoS Policy 876

68 CONGESTION MANAGEMENT

Overview 877
Congestion Management Policy 877
Configuring an SP Queue 879

Configuring a WRR Queue 880
Configuring SP+WRR Queues 881
Displaying and Maintaining Congestion Management 882

69 PRIORITY MAPPING

Priority Mapping Overview 883
Configuring a Priority Mapping Table 884
Configuring the Port Priority 885
Configuring Port Priority Trust Mode 886
Displaying and Maintaining Priority Mapping 887

70 APPLYING A QoS POLICY TO VLANS

Overview 889
Applying a QoS Policy to VLANs 889
Displaying and Maintaining QoS Policies Applied to VLANs 890
Configuration Examples 890

71 TRAFFIC MIRRORING CONFIGURATION

Overview 891
Configuring Traffic Mirroring 891
Displaying and Maintaining Traffic Mirroring 892
Traffic Mirroring Configuration Examples 892

72 PORT MIRRORING CONFIGURATION

Introduction to Port Mirroring 895
Configuring Local Port Mirroring 897
Configuring Remote Port Mirroring 898
Displaying and Maintaining Port Mirroring 899
Port Mirroring Configuration Examples 900

73 CLUSTER MANAGEMENT CONFIGURATION

Cluster Management Overview 905
Cluster Configuration Task List 911
Configuring the Management Device 912
Configuring the Member Devices 917
Configuring Access Between the Management Device and Its Member Devices 918
Adding a Candidate Device to a Cluster 919
Configuring Advanced Cluster Functions 919
Displaying and Maintaining Cluster Management 922
Cluster Management Configuration Examples 922

74 UDP HELPER CONFIGURATION

Introduction to UDP Helper 927
Configuring UDP Helper 927

Displaying and Maintaining UDP Helper 928
UDP Helper Configuration Example 928

75 SNMP CONFIGURATION

SNMP Overview 931
SNMP Configuration 933
Configuring SNMP Logging 935
Trap Configuration 936
Displaying and Maintaining SNMP 937
SNMP Configuration Example 938
SNMP Logging Configuration Example 939

76 RMON CONFIGURATION

RMON Overview 941
Configuring RMON 943
Displaying and Maintaining RMON 944
RMON Configuration Example 945

77 NTP CONFIGURATION

NTP Overview 947
NTP Configuration Task list 953
Configuring the Operation Modes of NTP 953
Configuring Optional Parameters of NTP 956
Configuring Access-Control Rights 957
Configuring NTP Authentication 958
Displaying and Maintaining NTP 960
NTP Configuration Examples 960

78 DNS CONFIGURATION

DNS Overview 971
Configuring the DNS Client 973
Configuring the DNS Proxy 974
Displaying and Maintaining DNS 974
DNS Configuration Examples 975
Troubleshooting DNS Configuration 980

79 FILE SYSTEM MANAGEMENT CONFIGURATION

File System Management 981
Configuration File Management 985
Displaying and Maintaining Device Configuration 989

80 FTP CONFIGURATION

FTP Overview 991
Configuring the FTP Client 992

Configuring the FTP Server 996
Displaying and Maintaining FTP 999

81 TFTP CONFIGURATION

TFTP Overview 1001
Configuring the TFTP Client 1002
Displaying and Maintaining the TFTP Client 1003
TFTP Client Configuration Example 1003

82 INFORMATION CENTER CONFIGURATION

Information Center Overview 1005
Configuring Information Center 1009
Displaying and Maintaining Information Center 1015
Information Center Configuration Examples 1015

83 BASIC CONFIGURATIONS

Basic Configurations 1021
CLI Features 1027

84 SYSTEM MAINTAINING AND DEBUGGING

System Maintaining and Debugging Overview 1033
System Maintaining and Debugging 1035
System Maintaining Example 1036

85 DEVICE MANAGEMENT

Device Management Overview 1039
Configuring Device Management 1039
Displaying and Maintaining Device Management Configuration 1043
Device Management Configuration Example 1043

86 NQA CONFIGURATION

NQA Overview 1047
NQA Configuration Task List 1050
Configuring the NQA Server 1050
Enabling the NQA Client 1051
Creating an NQA Test Group 1051
Configuring an NQA Test Group 1051
Configuring the Collaboration Function 1061
Configuring Trap Delivery 1061
Configuring Optional Parameters Common to an NQA Test Group 1062
Scheduling an NQA Test Group 1063
Displaying and Maintaining NQA 1064
NQA Configuration Examples 1064

87	VRRP CONFIGURATION	
	Introduction to VRRP	1073
	Configuring VRRP for IPv4	1081
	Configuring VRRP for IPv6	1084
	IPv4-Based VRRP Configuration Examples	1088
	IPv6-Based VRRP Configuration Examples	1096
	Troubleshooting VRRP	1105
88	SSH CONFIGURATION	
	SSH2.0 Overview	1107
	Configuring the Device as an SSH Server	1110
	Configuring the Device as an SSH Client	1115
	Displaying and Maintaining SSH	1118
	SSH Server Configuration Examples	1119
	SSH Client Configuration Examples	1125
89	SFTP SERVICE	
	SFTP Overview	1131
	Configuring an SFTP Server	1131
	Configuring an SFTP Client	1132
	SFTP Configuration Example	1135
90	RRPP CONFIGURATION	
	RRPP Overview	1139
	RRPP Configuration Task List	1146
	Configuring Master Node	1147
	Configuring Transit Node	1148
	Configuring Edge Node	1149
	Configuring Assistant Edge Node	1151
	Displaying and Maintaining RRPP	1152
	RRPP Typical Configuration Examples	1152
91	PORT SECURITY CONFIGURATION	
	Introduction to Port Security	1161
	Port Security Configuration Task List	1164
	Enabling Port Security	1164
	Setting the Maximum Number of Secure MAC Addresses	1165
	Setting the Port Security Mode	1165
	Configuring Port Security Features	1167
	Configuring Secure MAC Addresses	1168
	Ignoring the Authorization Information from the Server	1168
	Displaying and Maintaining Port Security	1169
	Port Security Configuration Examples	1169
	Troubleshooting Port Security	1178

92 LLDP CONFIGURATION

- Introduction to LLDP 1181
- LLDP Configuration Tasks List 1184
- Performing Basic LLDP Configuration 1184
- Configuring LLDP Trap 1188
- Displaying and Maintaining LLDP 1188
- LLDP Configuration Example 1189

93 POE CONFIGURATION

- PoE Overview 1193
- PoE Configuration Task List 1194
- Configuring the PoE Interface 1194
- Configuring PD Power Management 1196
- Configuring a Power Alarm Threshold for the PSE 1197
- Upgrading PSE Processing Software Online 1197
- Configuring a PD Disconnection Detection Mode 1198
- Enabling the PSE to Detect Nonstandard PDs 1198
- Displaying and Maintaining PoE 1199
- PoE Configuration Example 1199
- Troubleshooting PoE 1200

94 sFLOW CONFIGURATION

- sFlow Overview 1203
- Configuring sFlow 1204
- Displaying sFlow 1204
- sFlow Configuration Example 1204
- Troubleshooting sFlow Configuration 1206

95 SSL CONFIGURATION

- SSL Overview 1207
- SSL Configuration Task List 1208
- Configuring an SSL Server Policy 1208
- Configuring an SSL Client Policy 1210
- Displaying and Maintaining SSL 1211
- Troubleshooting SSL 1211

96 HTTPS CONFIGURATION

- HTTPS Overview 1213
- HTTPS Configuration Task List 1213
- Associating the HTTPS Service with an SSL Server Policy 1214
- Enabling the HTTPS Service 1214
- Associating the HTTPS Service with a Certificate Attribute Access Control Policy 1215
- Associating the HTTPS Service with an ACL 1215
- Displaying and Maintaining HTTPS 1215

HTTPS Configuration Example 1215

97 PKI CONFIGURATION

Introduction to PKI 1219
PKI Configuration Task List 1222
Configuring an Entity DN 1222
Configuring a PKI Domain 1223
Submitting a PKI Certificate Request 1225
Retrieving a Certificate Manually 1226
Configuring PKI Certificate Validation 1227
Destroying a Local RSA Key Pair 1228
Deleting a Certificate 1229
Configuring an Access Control Policy 1229
Displaying and Maintaining PKI 1229
PKI Configuration Examples 1230
Troubleshooting PKI 1235

98 TRACK CONFIGURATION

Track Overview 1237
Track Configuration Task List 1238
Configuring Collaboration Between the Track Module and the Detection Modules 1238
Configuring Collaboration Between the Track Module and the Application Modules 1239
Displaying and Maintaining Track Object(s) 1241
Track Configuration Example 1241

A ACRONYMS

ABOUT THIS GUIDE

This guide describes the 3Com® Switch 4800G and how to install hardware, configure and boot software, and maintain software and hardware. This guide also provides troubleshooting and support information for your switch.

This guide is intended for Qualified Service personnel who are responsible for configuring, using, and managing the switches. It assumes a working knowledge of local area network (LAN) operations and familiarity with communication protocols that are used to interconnect LANs.



Always download the Release Notes for your product from the 3Com World Wide Web site and check for the latest updates to software and product documentation:

<http://www.3com.com>

Conventions

Table 1 lists icon conventions that are used throughout this guide.

Table 1 Notice Icons

Icon	Notice Type	Description
	Information note	Information that describes important features or instructions.
	Caution	Information that alerts you to potential loss of data or potential damage to an application, system, or device.
	Warning	Information that alerts you to potential personal injury.

Related Documentation

The following manuals offer additional information necessary for managing your Switch Switch 4800G:

- *Switch 4800G Command Reference Guide* — Provides detailed descriptions of command line interface (CLI) commands, that you require to manage your Switch 4800G.
- *Switch 4800G Configuration Guide*— Describes how to configure your Switch 4800G using the supported protocols and CLI commands.
- *Switch 4800G Quick Reference Guides* — Provides a summary of command line interface (CLI) commands that are required for you to manage your Switch 4800G.

- *Switch 4800G Release Notes* — Contains the latest information about your product. If information in this guide differs from information in the release notes, use the information in the *Release Notes*.

These documents are available in Adobe Acrobat Reader Portable Document Format (PDF) on the CD-ROM that accompanies your router or on the 3Com World Wide Web site:

<http://www.3com.com/>

PRODUCT OVERVIEW

Preface

3Com Switch 4800G Family (hereinafter referred to as the Switch 4800G) are Gigabit Ethernet switching products developed by 3Com. The Switch 4800G have abundant service features. They provide the IPv6 forwarding function and 10GE uplink interfaces.

Through 3Com-specific cluster management, you can streamline network management. The Switch 4800G are designed as convergence and access devices for intranets and metropolitan area networks (MANs). Supporting IPv4/IPv6 dual-stack, the Switch 4800G provide abundant service features and routing functions and can also be used for connecting server groups in data centers.

Product Models

Table 2 Models in the 3Com Switch 4800G Family

Model	Number of service ports	Ports	Console port
3Com 3CRS48G-24-91	28	24 10/100/1,000 M electrical ports + 4 Gigabit SFP Combo ports + 2 10GE module slots	1
3Com 3CRS48G-48-91	52	48 10/100/1,000 M electrical ports + 4 Gigabit SFP Combo ports + 2 10GE module slots	1
3Com 3CRS48G-24P-91	28	24 10/100/1,000 M PoE electrical ports + 4 Gigabit SFP Combo ports + 2 10GE module slots	1
3Com 3CRS48G-48P-91	52	48 10/100/1,000 M PoE electrical ports + 4 Gigabit SFP Combo ports + 2 10GE module slots	1
3Com 3CRS48G-24S-91	28	24 100/1,000 M SFP ports + 8 10/100/1,000 M Combo electrical ports + 2 10GE module slots	1

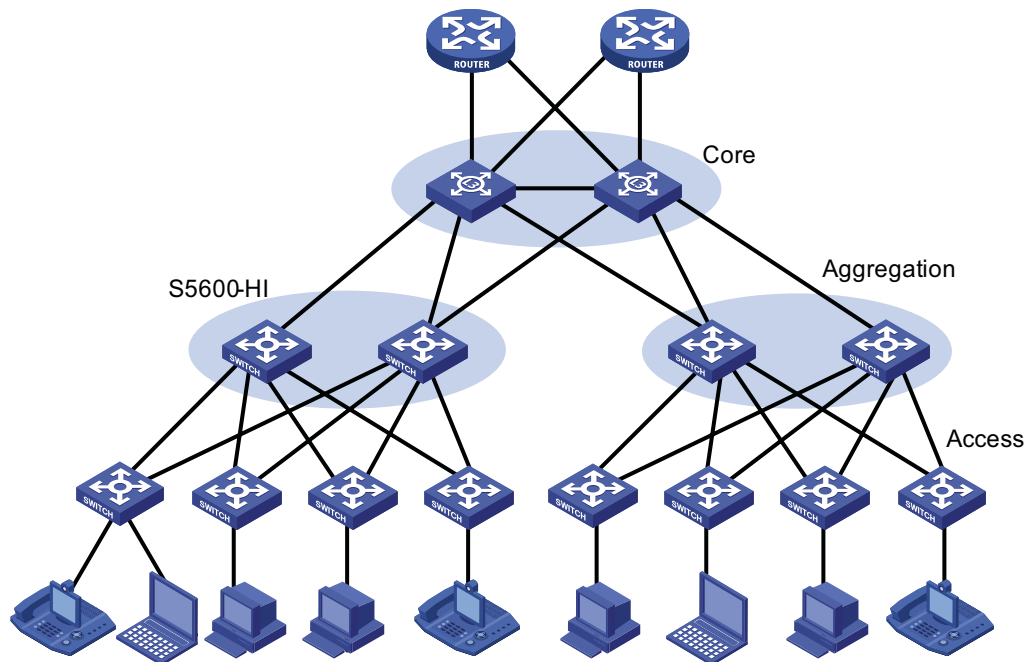
NETWORKING APPLICATIONS

The Switch 4800G are designed as convergence layer switches or access layer switches for enterprise networks and MANs. The Switch 4800G provide 24 or 48 autosensing Gigabit Ethernet ports and four SFP Combo Gigabit optical interfaces. In addition, the Switch 4800G provide two extension slots. You can configure XFP/CX4 extension module and up to four 10GE ports are supported. Networking is very flexible. The Switch 4800G can apply to Gigabit Ethernet to the desktop (GTTD) access of enterprise networks, user access of campus networks, and connection of data center server clusters. Several typical networking applications are described as follows.

Serving as a Convergence Layer Device

In medium- and large-sized enterprises or campus networks, the the Switch 4800G can serve as convergence layer switches that provide high-performance and large-capacity switching service and support 10GE uplink interfaces, which provide larger bandwidth for the devices.

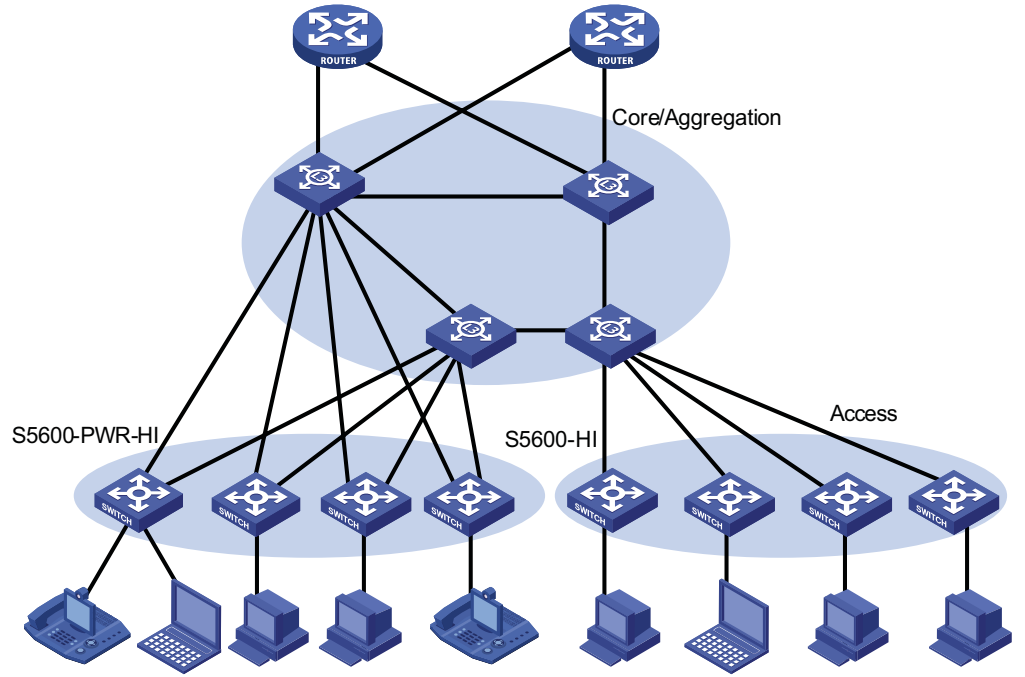
Figure 1 Application of the Switch 4800G at the convergence layer of enterprise networks/campus networks



Serving as a Access Layer Device

The Switch 4800G can serve as access layer switches that provide large access bandwidth and high port density. The Switch 4800G also provide PoE. Through Ethernet cables, the Switch 4800G can provide power to IP phone, WLAN AP, and other PD devices that support IEEE 802.3af to facilitate network maintenance and management.

Figure 2 Application of Switch 4800G at access layer



1

LOGGING IN TO AN ETHERNET SWITCH

Logging In to an Ethernet Switch

You can log in to an Switch 4800G in one of the following ways:

- Logging in locally through the console port
- Telnetting locally or remotely to an Ethernet port
- Telnetting to the console port using a modem
- Logging in to the Web-based network management system
- Logging in through NMS (network management station)

Introduction to the User Interface

Supported User Interfaces

Switch 4800G supports two types of user interfaces: AUX and VTY.

Table 3 Description on user interface

User interface	Applicable user	Port used	Description
AUX	Users logging in through the console port	Console port	Each switch can accommodate one AUX user.
VTY	Telnet users and SSH users	Ethernet port	Each switch can accommodate up to five VTY users.



As the AUX port and the console port of a 3Com switch are the same one, you will be in the AUX user interface if you log in through this port.

User Interface Number

Two kinds of user interface index exist: absolute user interface index and relative user interface index.

- 1 The absolute user interface indexes are as follows:
 - AUX user interface: 0
 - VTY user interfaces: Numbered after AUX user interfaces and increases in the step of 1
- 2 A relative user interface index can be obtained by appending a number to the identifier of a user interface type. It is generated by user interface type. The relative user interface indexes are as follows:
 - AUX user interface: AUX 0
 - VTY user interfaces: VTY 0, VTY 1, VTY 2, and so on.

Common User Interface Configuration

To do...	Use the command...	Remarks
Lock the current user interface	lock	Optional Execute this command in user view. A user interface is not locked by default.
Specify to send messages to all user interfaces/a specified user interface	send { all <i>number</i> <i>type number</i> }	Optional Execute this command in user view.
Disconnect a specified user interface	free user-interface [<i>type</i>] <i>number</i>	Optional Execute this command in user view.
Enter system view	system-view	-
Set the banner	header { incoming legal login shell motd } <i>text</i>	Optional
Set a system name for the switch	sysname <i>string</i>	Optional
Enter user interface view	user-interface [<i>type</i>] <i>first-number</i> [<i>last-number</i>]	-
Define a shortcut key for aborting tasks	escape-key { default <i>character</i> }	Optional The default shortcut key combination for aborting tasks is < Ctrl+C >.
Set the history command buffer size	history-command max-size <i>value</i>	Optional The default history command buffer size is 10. That is, a history command buffer can store up to 10 commands by default.
Set the timeout time for the user interface	idle-timeout <i>minutes</i> [<i>seconds</i>]	Optional The default timeout time of a user interface is 10 minutes. With the timeout time being 10 minutes, the connection to a user interface is terminated if no operation is performed in the user interface within 10 minutes. You can use the idle-timeout 0 command to disable the timeout function.
Set the maximum number of lines the screen can contain	screen-length <i>screen-length</i>	Optional By default, the screen can contain up to 24 lines. You can use the screen-length 0 command to disable the function to display information in pages.
Make terminal services available	shell	Optional By default, terminal services are available in all user interfaces.

To do...	Use the command...	Remarks
Set the display type of a terminal	terminal type { ansi vt100 }	Optional By default, the terminal display type is ANSI. The device must use the same type of display as the terminal. If the terminal uses VT 100, the device should also use VT 100.
Display the information about the current user interface/all user interfaces	display users [all]	You can execute this command in any view.
Display the physical attributes and configuration of the current/a specified user interface	display user-interface [<i>type number</i> <i>number</i>] [summary]	You can execute this command in any view.
Display the information about the current web users	display web users	You can execute this command in any view.

2

LOGGING IN THROUGH THE CONSOLE PORT



The default system name of the Switch 4800G is 3Com, that is, the command line prompt is 3Com. All the following examples take 3Com as the command line prompt.

Introduction

To log in through the console port is the most common way to log in to a switch. It is also the prerequisite to configure other login methods. By default, you can log in to an Switch 4800G through its console port only.

To log in to an Ethernet switch through its console port, the related configuration of the user terminal must be in accordance with that of the console port.

Table 4 lists the default settings of a console port.

Table 4 The default settings of a console port

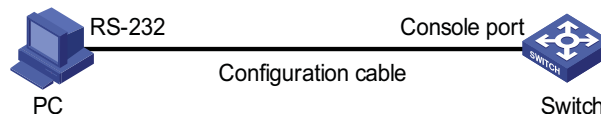
Setting	Default
Baud rate	9,600 bps
Flow control	Off
Check mode	No check bit
Stop bits	1
Data bits	8

After logging in to a switch, you can perform configuration for AUX users. Refer to section "Console Port Login Configuration" on page 33 for more.

Setting Up the Connection to the Console Port

- Connect the serial port of your PC/terminal to the console port of the switch, as shown in Figure 3.

Figure 3 Diagram for setting the connection to the console port



- If you use a PC to connect to the console port, launch a terminal emulation utility (such as Terminal in Windows 3.X or HyperTerminal in Windows 9X/Windows 2000/Windows XP) and perform the configuration shown in Figure 4 through Figure 6 for the connection to be created. Normally, the parameters of a terminal are configured as those listed in Table 4.

Figure 4 Create a connection

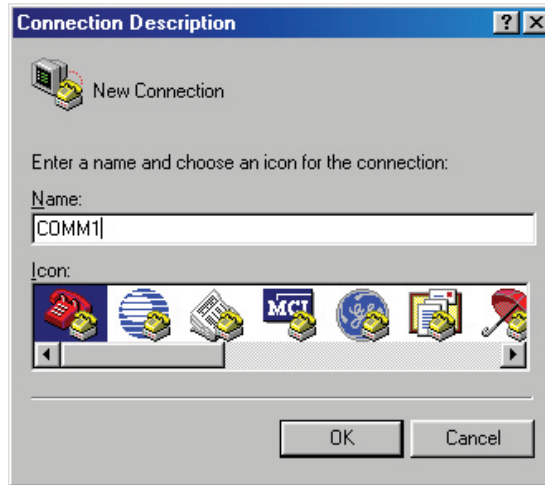
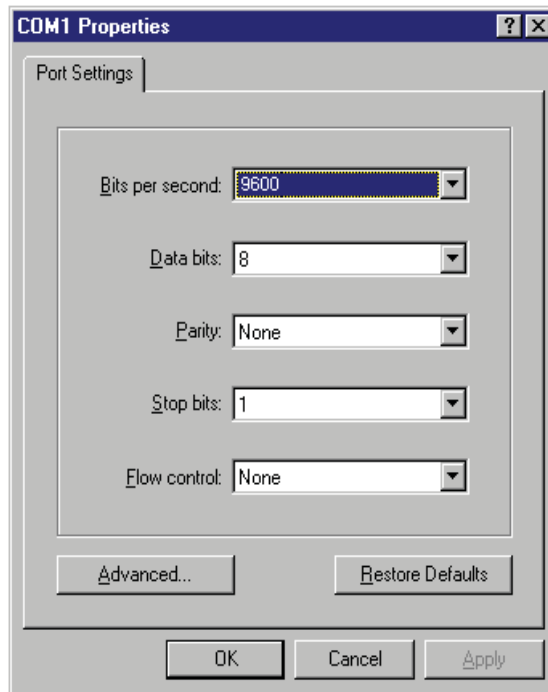


Figure 5 Specify the port used to establish the connection



Figure 6 Set port parameters terminal window



- Turn on the switch. The user will be prompted to press the Enter key if the switch successfully completes POST (power-on self test). The prompt (such as <SW4800G>) appears after the user presses the Enter key.
- You can then configure the switch or check the information about the switch by executing commands. You can also acquire help by type the ? character. Refer to the following chapters for information about the commands.

Console Port Login Configuration

Common Configuration Table 5 lists the common configuration of console port login.

Table 5 Common configuration of console port login

Configuration	Description
Console port configuration	Optional
Baud rate	The default baud rate is 9,600 bps.
Check mode	Optional
	By default, the check mode of the console port is set to "none", which means no check bit.
Stop bits	Optional
	The default stop bits of a console port is 1.
Data bits	Optional
	The default data bits of a console port is 8.

Table 5 Common configuration of console port login

Configuration		Description
AUX user interface configuration	Configure the command level available to the users logging in to the AUX user interface	Optional By default, commands of level 3 are available to the users logging in to the AUX user interface.
Terminal configuration	Define a shortcut key for aborting tasks	Optional The default shortcut key combination for aborting tasks is < Ctrl + C >.
	Define a shortcut key for starting terminal sessions	Optional By default, pressing Enter key starts the terminal session.
	Make terminal services available	Optional By default, terminal services are available in all user interfaces
	Set the maximum number of lines the screen can contain	Optional By default, the screen can contain up to 24 lines.
	Set history command buffer size	Optional By default, the history command buffer can contain up to 10 commands.
	Set the timeout time of a user interface	Optional The default timeout time is 10 minutes.



CAUTION: Changing of console port configuration terminates the connection to the console port. To establish the connection again, you need to modify the configuration of the termination emulation utility running on your PC accordingly. Refer to section "Setting Up the Connection to the Console Port" on page 31 for more.

Console Port Login Configurations for Different Authentication Modes

Table 6 lists console port login configurations for different authentication modes.

Table 6 Console port login configurations for different authentication modes

Authentication mode	Console port login configuration	Description
None	Perform common configuration	Optional Refer to section "Common Configuration" on page 33 for more.
	Perform common configuration	Optional Refer to section "Common Configuration" on page 33 for more.
Password	Configure the password	Required
	Perform common configuration	Optional Refer to section "Common Configuration" on page 33 for more.

Table 6 Console port login configurations for different authentication modes

Authentication mode	Console port login configuration		Description
Scheme	Specify to perform local authentication or RADIUS authentication	AAA configuration specifies whether to perform local authentication or RADIUS authentication	Optional Local authentication is performed by default. Refer to "AAA/RADIUS/HWTACACS Overview" on page 747.
	Configure user name and password	Configure user names and passwords for local/remote users	Required <ul style="list-style-type: none"> ■ The user name and password of a local user are configured on the switch. ■ The user name and password of a remote user are configured on the RADIUS server. Refer to "Configuring RADIUS" on page 765.
	Manage AUX users	Set service type for AUX users	Required
	Perform common configuration	Perform common configuration for console port login	Optional Refer to section "Common Configuration" on page 33 for more.



Changes of the authentication mode of console port login will not take effect unless you exit and enter again the CLI.

Console Port Login Configuration with Authentication Mode Being None

Configuration Procedure

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter AUX user interface view	user-interface aux 0	-
Configure not to authenticate users	authentication-mode none	Required By default, users logging in through the console port are not authenticated.

To do...	Use the command...	Remarks
Configure the console port	Set the baud rate	speed <i>speed-value</i> Optional The default baud rate of an AUX port (also the console port) is 9,600 bps.
	Set the check mode	parity { even mark none odd space } Optional By default, the check mode of a console port is set to none , that is, no check bit.
	Set the stop bits	stopbits { 1 1.5 2 } Optional The stop bits of a console port is 1.
	Set the data bits	databits { 5 6 7 8 } Optional The default data bits of a console port is 8.
Configure the command level available to users logging in to the user interface	user privilege level <i>level</i> Optional By default, commands of level 3 are available to users logging in to the AUX user interface.	
Define a shortcut key for starting terminal sessions	activation-key <i>character</i> Optional By default, pressing Enter key starts the terminal session.	
Define a shortcut key for aborting tasks	escape-key { default <i>character</i> } Optional The default shortcut key combination for aborting tasks is < Ctrl+C >.	
Make terminal services available	shell Optional By default, terminal services are available in all user interfaces.	
Set the maximum number of lines the screen can contain	screen-length <i>screen-length</i> Optional By default, the screen can contain up to 24 lines. You can use the screen-length 0 command to disable the function to display information in pages.	
Set the history command buffer size	history-command max-size <i>value</i> Optional The default history command buffer size is 10. That is, a history command buffer can store up to 10 commands by default.	
Set the timeout time for the user interface	idle-timeout <i>minutes</i> [<i>seconds</i>] Optional The default timeout time of a user interface is 10 minutes. With the timeout time being 10 minutes, the connection to a user interface is terminated if no operation is performed in the user interface within 10 minutes. You can use the idle-timeout 0 command to disable the timeout function.	

Note that if you configure not to authenticate the users, the command level available to users logging in to a switch depends on both the **authentication-mode none** command and the **user privilege level *level*** command, as listed in the following table.

Table 7 Determine the command level (A)

Scenario			
Authentication mode	User type	Command	Command level
None (authentication-mode none)	Users logging in through console ports	The user privilege level <i>level</i> command not executed	Level 3
		The user privilege level <i>level</i> command already executed	Determined by the <i>level</i> argument

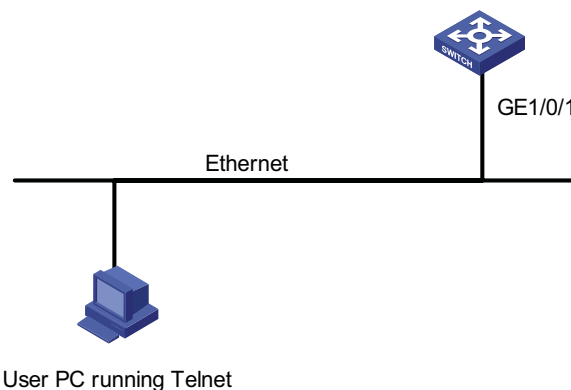
Configuration Example Network requirements

Assume the switch is configured to allow you to login through Telnet, and your user level is set to the administrator level (level 3). After you telnet to the switch, you need to limit the console user at the following aspects.

- The user is not authenticated when logging in through the console port.
- Commands of level 2 are available to user logging in to the AUX user interface.
- The baud rate of the console port is 19,200 bps.
- The screen can contain up to 30 lines.
- The history command buffer can contain up to 20 commands.
- The timeout time of the AUX user interface is 6 minutes.

Network diagram

Figure 7 Network diagram for AUX user interface configuration (with the authentication mode being none)



Configuration procedure

Enter system view.

```
<SW4800G> system-view
```

Enter AUX user interface view.

```
[SW4800G] user-interface aux 0

# Specify not to authenticate the user logging in through the console port.

[SW4800G-ui-aux0] authentication-mode none

# Specify commands of level 2 are available to the user logging in to the AUX user
interface.

[SW4800G-ui-aux0] user privilege level 2

# Set the baud rate of the console port to 19,200 bps.

[SW4800G-ui-aux0] speed 19200

# Set the maximum number of lines the screen can contain to 30.

[SW4800G-ui-aux0] screen-length 30

# Set the maximum number of commands the history command buffer can store
to 20.

[SW4800G-ui-aux0] history-command max-size 20

# Set the timeout time of the AUX user interface to 6 minutes.

[SW4800G-ui-aux0] idle-timeout 6
```

After the above configuration, to ensure a successful login, the console user needs to change the corresponding configuration of the terminal emulation program running on the PC, to make the configuration consistent with that on the switch. Refer to section “Setting Up the Connection to the Console Port” on page 31 for more.

Console Port Login Configuration with Authentication Mode Being Password

Configuration Procedure

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter AUX user interface view	user-interface aux 0	-
Configure to authenticate users using the local password	authentication-mode password	Required By default, users logging in through the console port are not authenticated, while users logging in through the Modem or Telnet need to pass the password authentication.
Set the local password	set authentication password { cipher simple } password	Required

To do...	Use the command...	Remarks
Configure the console port	Set the baud rate	speed <i>speed-value</i> Optional The default baud rate of an AUX port (also the console port) is 9,600 bps.
	Set the check mode	parity { even mark none odd space } Optional By default, the check mode of a console port is set to none , that is, no check bit.
	Set the stop bits	stopbits { 1 1.5 2 } Optional The default stop bits of a console port is 1.
	Set the data bits	databits { 5 6 7 8 } Optional The default data bits of a console port is 8.
Configure the command level available to users logging in to the user interface	user privilege level <i>level</i> Optional By default, commands of level 3 are available to users logging in to the AUX user interface.	
Define a shortcut key for starting terminal sessions	activation-key <i>character</i> Optional By default, pressing Enter key starts the terminal session.	
Define a shortcut key for aborting tasks	escape-key { default <i>character</i> } Optional The default shortcut key combination for aborting tasks is < Ctrl+C >.	
Make terminal services available to the user interface	shell Optional By default, terminal services are available in all user interfaces.	
Set the maximum number of lines the screen can contain	screen-length <i>screen-length</i> Optional By default, the screen can contain up to 24 lines. You can use the screen-length 0 command to disable the function to display information in pages.	
Set history command buffer size	history-command max-size <i>value</i> Optional The default history command buffer size is 10. That is, a history command buffer can store up to 10 commands by default.	
Set the timeout time for the user interface	idle-timeout <i>minutes</i> [<i>seconds</i>] Optional The default timeout time of a user interface is 10 minutes. With the timeout time being 10 minutes, the connection to a user interface is terminated if no operation is performed in the user interface within 10 minutes. You can use the idle-timeout 0 command to disable the timeout function.	

Note that if you configure to authenticate the users in the password mode, the command level available to users logging in to a switch depends on both the

authentication-mode password and the **user privilege level** *level* command, as listed in the following table.

Table 8 Determine the command level (B)

Scenario			
Authentication mode	User type	Command	Command level
Local authentication (authentication-mode password)	Users logging in to the AUX user interface	The user privilege level <i>level</i> command not executed	Level 3
		The user privilege level <i>level</i> command already executed	Determined by the <i>level</i> argument

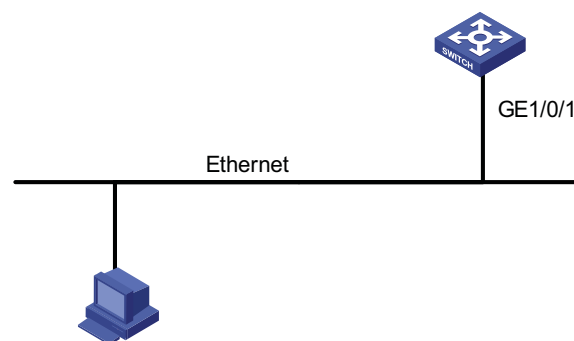
Configuration Example Network requirements

Assume the switch is configured to allow you to login through Telnet, and your user level is set to the administrator level (level 3). After you telnet to the switch, you need to limit the console user at the following aspects.

- The user is authenticated against the local password when logging in through the console port.
- The local password is set to 123456 (in plain text).
- The commands of level 2 are available to users logging in to the AUX user interface.
- The baud rate of the console port is 19,200 bps.
- The screen can contain up to 30 lines.
- The history command buffer can store up to 20 commands.
- The timeout time of the AUX user interface is 6 minutes.

Network diagram

Figure 8 Network diagram for AUX user interface configuration (with the authentication mode being password)



User PC running Telnet

Configuration procedure

Enter system view.

```
<SW4800G> system-view
```

Enter AUX user interface view.

```
[SW4800G] user-interface aux 0

# Specify to authenticate the user logging in through the console port using the
local password.

[SW4800G-ui-aux0] authentication-mode password

# Set the local password to 123456 (in plain text).

[SW4800G-ui-aux0] set authentication password simple 123456

# Specify commands of level 2 are available to the user logging in to the AUX user
interface.

[SW4800G-ui-aux0] user privilege level 2

# Set the baud rate of the console port to 19,200 bps.

[SW4800G-ui-aux0] speed 19200

# Set the maximum number of lines the screen can contain to 30.

[SW4800G-ui-aux0] screen-length 30

# Set the maximum number of commands the history command buffer can store
to 20.

[SW4800G-ui-aux0] history-command max-size 20

# Set the timeout time of the AUX user interface to 6 minutes.

[SW4800G-ui-aux0] idle-timeout 6
```

After the above configuration, to ensure a successful login, the console user needs to change the corresponding configuration of the terminal emulation program running on the PC, to make the configuration consistent with that on the switch. Refer to section "Setting Up the Connection to the Console Port" on page 31 for more.

Console Port Login Configuration with Authentication Mode Being Scheme

Configuration Procedure

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...		Use the command...	Remarks
Configure the authentication mode	Enter the default ISP domain view	domain <i>Domain name</i>	Optional
	Specify the AAA scheme to be applied to the domain	authentication default { hwtacacs- scheme <i>hwtacacs-scheme-name</i> [local] local none radius-scheme <i>radius-scheme-name</i> [local] }	By default, the local AAA scheme is applied. If you specify to apply the local AAA scheme, you need to perform the configuration concerning local user as well.
	Quit to system view	quit	If you specify to apply an existing scheme by providing the <i>radius-scheme-name</i> argument, you need to perform the following configuration as well: <ul style="list-style-type: none"> ■ Perform AAA-RADIUS configuration on the switch. (Refer to “AAA/RADIUS/HWTA CACS Configuration” on page 747 for more.) ■ Configure the user name and password accordingly on the AAA server. (Refer to “Configuring AAA” on page 758.)
	Create a local user (Enter local user view.)	local-user <i>user-name</i>	Required No local user exists by default.
	Set the authentication password for the local user	password { simple cipher } <i>password</i>	Required
	Specify the service type for AUX users	service-type terminal [level <i>level</i>]	Required
	Quit to system view	quit	-
	Enter AUX user interface view	user-interface aux 0	-
	Configure to authenticate users locally or remotely	authentication-mode scheme [command-authorization]	Required The specified AAA scheme determines whether to authenticate users locally or remotely. Users are authenticated locally by default.

To do...	Use the command...	Remarks
Configure the console port	speed <i>speed-value</i>	Optional The default baud rate of the AUX port (also the console port) is 9,600 bps.
	parity { even mark none odd space }	Optional By default, the check mode of the console port is set to none , that is, no check bit.
	stopbits { 1 1.5 2 }	Optional The default stop bits of the console port is 1.
	databits { 5 6 7 8 }	Optional The default data bits of the console port is 8.
Configure the command level available to users logging in to the user interface	user privilege level <i>level</i>	Optional By default, commands of level 3 are available to users logging in to the AUX user interface.
Define a shortcut key for starting terminal sessions	activation-key <i>character</i>	Optional By default, pressing Enter key starts the terminal session.
Define a shortcut key for aborting tasks	escape-key { default <i>character</i> }	Optional The default shortcut key combination for aborting tasks is < Ctrl+C >.
Make terminal services available to the user interface	shell	Optional By default, terminal services are available in all user interfaces.
Set the maximum number of lines the screen can contain	screen-length <i>screen-length</i>	Optional By default, the screen can contain up to 24 lines. You can use the screen-length 0 command to disable the function to display information in pages.
Set history command buffer size	history-command max-size <i>value</i>	Optional The default history command buffer size is 10. That is, a history command buffer can store up to 10 commands by default.

To do...	Use the command...	Remarks
Set the timeout time for the user interface	idle-timeout <i>minutes</i> [<i>seconds</i>]	Optional The default timeout time of a user interface is 10 minutes. With the timeout time being 10 minutes, the connection to a user interface is terminated if no operation is performed in the user interface within 10 minutes. You can use the idle-timeout 0 command to disable the timeout function.

Note that the level the commands of which are available to users logging in to a switch depends on the **authentication-mode scheme** [**command-authorization**] command, the **user privilege level** *level* command, and the **service-type terminal** [**level** *level*] command, as listed in Table 9.

Table 9 Determine the command level

Scenario			
Authentication mode	User type	Command	Command level
authentication-mode scheme [command-authorization]	Users logging in to the console port and pass AAA-RADIUS or local authentication	The user privilege level <i>level</i> command is not executed, and the service-type terminal [level <i>level</i>] command does not specify the available command level.	Level 0 The default command level available for local users is level 0.
		The user privilege level <i>level</i> command is not executed, and the service-type terminal [level <i>level</i>] command specifies the available command level.	Determined by the service-type terminal [level <i>level</i>] command
		The user privilege level <i>level</i> command is executed, and the service-type terminal [level <i>level</i>] command does not specify the available command level.	Level 0
		The user privilege level <i>level</i> command is executed, and the service-type terminal [level <i>level</i>] command specifies the available command level.	Determined by the service-type terminal [level <i>level</i>] command

Configuration Example Network requirements

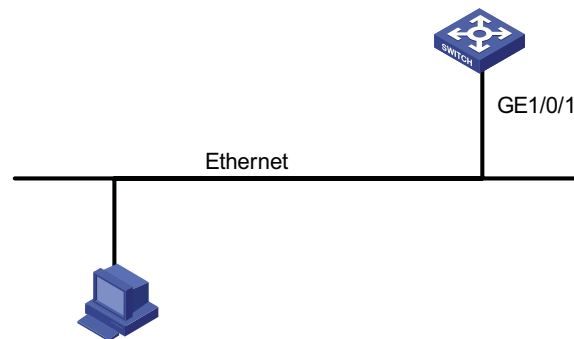
Assume the switch is configured to allow you to login through Telnet, and your user level is set to the administrator level (level 3). After you telnet to the switch, you need to limit the console user at the following aspects.

- Configure the name of the local user to be **guest**.
- Set the authentication password of the local user to **123456** (in plain text).

- Set the service type of the local user to Terminal.
- Configure to authenticate the user logging in through the console port in the scheme mode.
- The commands of level 2 are available to the user logging in to the AUX user interface.
- The baud rate of the console port is 19,200 bps.
- The screen can contain up to 30 lines.
- The history command buffer can store up to 20 commands.
- The timeout time of the AUX user interface is 6 minutes.

Network diagram

Figure 9 Network diagram for AUX user interface configuration (with the authentication mode being scheme)



User PC running Telnet

Configuration procedure

Enter system view.

```
<SW4800G> system-view
```

Create a local user named guest and enter local user view.

```
[SW4800G] local-user guest
```

Set the authentication password to **123456** (in plain text).

```
[SW4800G-luser-guest] password simple 123456
```

Set the service type to Terminal, Specify commands of level 2 are available to the user logging in to the AUX user interface.

```
[SW4800G-luser-guest] service-type terminal level 2
```

```
[SW4800G-luser-guest] quit
```

Enter AUX user interface view.

```
[SW4800G] user-interface aux 0
```

Configure to authenticate the user logging in through the console port in the scheme mode.

```
[SW4800G-ui-aux0] authentication-mode scheme

# Set the baud rate of the console port to 19,200 bps.

[SW4800G-ui-aux0] speed 19200

# Set the maximum number of lines the screen can contain to 30.

[SW4800G-ui-aux0] screen-length 30

# Set the maximum number of commands the history command buffer can store
to 20.

[SW4800G-ui-aux0] history-command max-size 20

# Set the timeout time of the AUX user interface to 6 minutes.

[SW4800G-ui-aux0] idle-timeout 6
```

After the above configuration, to ensure a successful login, the console user needs to change the corresponding configuration of the terminal emulation program running on the PC, to make the configuration consistent with that on the switch. Refer to section "Setting Up the Connection to the Console Port" on page 31 for more.

3

LOGGING IN THROUGH TELNET

Introduction

You can telnet to a remote switch to manage and maintain the switch. To achieve this, you need to configure both the switch and the Telnet terminal properly.

Table 10 Requirements for Telnet to a switch

Item	Requirement
Switch	Start the Telnet Server The IP address of the VLAN of the switch is configured and the route between the switch and the Telnet terminal is available. (Refer to "IP Addressing Configuration" on page 121 and "IP Routing Overview" on page 241 for more.) The authentication mode and other settings are configured. Refer to Table 11 and Table 12.
Telnet terminal	Telnet is running. The IP address of the management VLAN of the switch is available.



- *After you log in to the switch through Telnet, you can issue commands to the switch by way of pasting session text, which cannot exceed 2000 bytes, and the pasted commands must be in the same view; otherwise, the switch may not execute the commands correctly.*
- *If the session text exceeds 2000 bytes, you can save it in a configuration file, upload the configuration file to the switch and reboot the switch with this configuration file. For details, refer to "Configuration File Management" on page 985.*
- *The way to log in to a switch using Telnet based on IPv6 is the same as that based on IPv4.*

Common Configuration

Table 11 lists the common Telnet configuration.

Table 11 Common Telnet configuration

Configuration	Description	
VTY user interface configuration	Configure the command level available to users logging in to the VTY user interface	Optional By default, commands of level 0 are available to users logging in to a VTY user interface.
	Configure the protocols the user interface supports	Optional By default, Telnet and SSH protocol are supported.
	Set the command that is automatically executed when a user logs into the user interface	Optional By default, no command is automatically executed when a user logs into a user interface.
VTY terminal configuration	Define a shortcut key for aborting tasks	Optional The default shortcut key combination for aborting tasks is < Ctrl+C >.
	Make terminal services available	Optional By default, terminal services are available in all user interfaces
	Set the maximum number of lines the screen can contain	Optional By default, the screen can contain up to 24 lines.
	Set history command buffer size	Optional By default, the history command buffer can contain up to 10 commands.
	Set the timeout time of a user interface	Optional The default timeout time is 10 minutes.

**CAUTION:**

- The **auto-execute command** command may cause you unable to perform common configuration in the user interface, so use it with caution.
- Before executing the **auto-execute command** command and save your configuration, make sure you can log in to the switch in other modes and cancel the configuration.

Telnet Configurations for Different Authentication Modes

Table 12 lists Telnet configurations for different authentication modes.

Table 12 Telnet configurations for different authentication modes

Authentication mode	Telnet configuration	Description
None	Perform common configuration	Perform common Telnet configuration Optional Refer to Table 11.
Password	Configure the password	Configure the password for local authentication Required
	Perform common configuration	Perform common Telnet configuration Optional Refer to Table 11.

Table 12 Telnet configurations for different authentication modes

Authentication mode	Telnet configuration		Description
Scheme	Specify to perform local authentication or RADIUS authentication	AAA configuration specifies whether to perform local authentication or RADIUS authentication	Optional Local authentication is performed by default. Refer to "AAA/RADIUS/HWTACACS Configuration" on page 747.
	Configure user name and password	Configure user names and passwords for local/remote users	Required <ul style="list-style-type: none"> The user name and password of a local user are configured on the switch. The user name and password of a remote user are configured on the RADIUS server. Refer to "Configuring RADIUS" on page 765 for more.
	Manage VTY users	Set service type for VTY users	Required
	Perform common configuration	Perform common Telnet configuration	Optional Refer to Table 11.

Telnet Configuration with Authentication Mode Being None

Configuration Procedure

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the Telnet server function	telnet server enable	Required
Enter one or more VTY user interface views	user-interface vty <i>first-number</i> [<i>last-number</i>]	-
Configure not to authenticate users logging in to VTY user interfaces	authentication-mode none	Required By default, VTY users are authenticated after logging in.
Configure the command level available to users logging in to VTY user interface	user privilege level <i>level</i>	Optional By default, commands of level 0 are available to users logging in to VTY user interfaces.
Configure the protocols to be supported by the VTY user interface	protocol inbound { all ssh telnet }	Optional By default, both Telnet protocol and SSH protocol are supported.
Set the command that is automatically executed when a user logs into the user interface	auto-execute command <i>text</i>	Optional By default, no command is automatically executed when a user logs into a user interface.

To do...	Use the command...	Remarks
Define a shortcut key for aborting tasks	escape-key { default <i>character</i> }	Optional The default shortcut key combination for aborting tasks is < Ctrl+C >.
Make terminal services available	shell	Optional By default, terminal services are available in all user interfaces.
Set the maximum number of lines the screen can contain	screen-length <i>screen-length</i>	Optional By default, the screen can contain up to 24 lines. You can use the screen-length 0 command to disable the function to display information in pages.
Set the history command buffer size	history-command max-size <i>value</i>	Optional The default history command buffer size is 10. That is, a history command buffer can store up to 10 commands by default.
Set the timeout time of the VTY user interface	idle-timeout <i>minutes</i> [<i>seconds</i>]	Optional The default timeout time of a user interface is 10 minutes. With the timeout time being 10 minutes, the connection to a user interface is terminated if no operation is performed in the user interface within 10 minutes. You can use the idle-timeout 0 command to disable the timeout function.

Note that if you configure not to authenticate the users, the command level available to users logging in to a switch depends on both the **authentication-mode none** command and the **user privilege level** *level* command, as listed in Table 13.

Table 13 Determine the command level when users logging in to switches are not authenticated

Scenario		Command	Command level
Authentication mode	User type		
None (authentication-mode none)	VTY users	The user privilege level <i>level</i> command not executed	Level 0
		The user privilege level <i>level</i> command already executed	Determined by the <i>level</i> argument

Configuration Example Network requirements

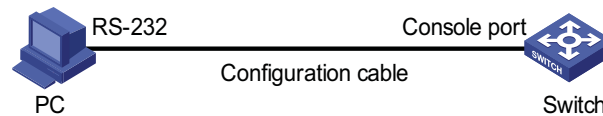
Assume that you are a level 3 AUX user and want to perform the following configuration for Telnet users logging in to VTY 0:

- Do not authenticate users logging in to VTY 0.

- Commands of level 2 are available to users logging in to VTY 0.
- Telnet protocol is supported.
- The screen can contain up to 30 lines.
- The history command buffer can contain up to 20 commands.
- The timeout time of VTY 0 is 6 minutes.

Network diagram

Figure 10 Network diagram for Telnet configuration (with the authentication mode being none)



Configuration procedure

Enter system view, and enable the Telnet service.

```
<SW4800G> system-view
[SW4800G] telnet server enable
```

Enter VTY 0 user interface view.

```
[SW4800G] user-interface vty 0
```

Configure not to authenticate Telnet users logging in to VTY 0.

```
[SW4800G-ui-vty0] authentication-mode none
```

Specify commands of level 2 are available to users logging in to VTY 0.

```
[SW4800G-ui-vty0] user privilege level 2
```

Configure Telnet protocol is supported.

```
[SW4800G-ui-vty0] protocol inbound telnet
```

Set the maximum number of lines the screen can contain to 30.

```
[SW4800G-ui-vty0] screen-length 30
```

Set the maximum number of commands the history command buffer can store to 20.

```
[SW4800G-ui-vty0] history-command max-size 20
```

Set the timeout time to 6 minutes.

```
[SW4800G-ui-vty0] idle-timeout 6
```

Telnet Configuration with Authentication Mode Being Password

Configuration Procedure

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the Telnet server function	telnet server enable	Required
Enter one or more VTY user interface views	user-interface vty <i>first-number</i> [<i>last-number</i>]	-
Configure to authenticate users logging in to VTY user interfaces using the local password	authentication-mode password	Required
Set the local password	set authentication password { cipher simple } password	Required
Configure the command level available to users logging in to the user interface	user privilege level <i>level</i>	Optional By default, commands of level 0 are available to users logging in to VTY user interface.
Configure the protocol to be supported by the user interface	protocol inbound { all ssh telnet }	Optional By default, both Telnet protocol and SSH protocol are supported.
Set the command that is automatically executed when a user logs into the user interface	auto-execute command <i>text</i>	Optional By default, no command is automatically executed when a user logs into a user interface.
Define a shortcut key for aborting tasks	escape-key { default character }	Optional The default shortcut key combination for aborting tasks is < Ctrl+C >.
Make terminal services available	shell	Optional By default, terminal services are available in all user interfaces.
Set the maximum number of lines the screen can contain	screen-length <i>screen-length</i>	Optional By default, the screen can contain up to 24 lines. You can use the screen-length 0 command to disable the function to display information in pages.
Set the history command buffer size	history-command max-size <i>value</i>	Optional The default history command buffer size is 10. That is, a history command buffer can store up to 10 commands by default.

To do...	Use the command...	Remarks
Set the timeout time of the user interface	idle-timeout <i>minutes</i> [<i>seconds</i>]	Optional The default timeout time of a user interface is 10 minutes. With the timeout time being 10 minutes, the connection to a user interface is terminated if no operation is performed in the user interface within 10 minutes. You can use the idle-timeout 0 command to disable the timeout function.

Note that if you configure to authenticate the users in the password mode, the command level available to users logging in to a switch depends on both the **authentication-mode password** command and the **user privilege level /level** command, as listed in Table 14.

Table 14 Determine the command level when users logging in to switches are authenticated in the password mode

Scenario			
Authentication mode	User type	Command	Command level
Password (authentication-mode password)	VTY users	The user privilege level /level command not executed	Level 0
		The user privilege level /level command already executed	Determined by the <i>level</i> argument

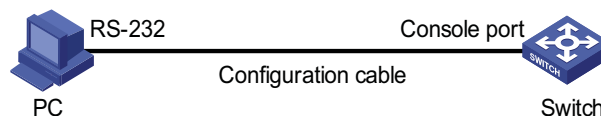
Configuration Example Network requirements

Assume that you are a level 3 AUX user and want to perform the following configuration for Telnet users logging in to VTY 0:

- Authenticate users logging in to VTY 0 using the local password.
- Set the local password to 123456 (in plain text).
- Commands of level 2 are available to users logging in to VTY 0.
- Telnet protocol is supported.
- The screen can contain up to 30 lines.
- The history command buffer can contain up to 20 commands.
- The timeout time of VTY 0 is 6 minutes.

Network diagram

Figure 11 Network diagram for Telnet configuration (with the authentication mode being password)



Configuration procedure

Enter system view, and enable the Telnet service.

```
<SW4800G> system-view
[SW4800G] telnet server enable
```

Enter VTY 0 user interface view.

```
[SW4800G] user-interface vty 0
```

Configure to authenticate users logging in to VTY 0 using the local password.

```
[SW4800G-ui-vty0] authentication-mode password
```

Set the local password to **123456** (in plain text).

```
[SW4800G-ui-vty0] set authentication password simple 123456
```

Specify commands of level 2 are available to users logging in to VTY 0.

```
[SW4800G-ui-vty0] user privilege level 2
```

Configure Telnet protocol is supported.

```
[SW4800G-ui-vty0] protocol inbound telnet
```

Set the maximum number of lines the screen can contain to 30.

```
[SW4800G-ui-vty0] screen-length 30
```

Set the maximum number of commands the history command buffer can store to 20.

```
[SW4800G-ui-vty0] history-command max-size 20
```

Set the timeout time to 6 minutes.

```
[SW4800G-ui-vty0] idle-timeout 6
```

Telnet Configuration with Authentication Mode Being Scheme

Configuration Procedure

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the Telnet server function	telnet server enable	Required

To do...	Use the command...	Remarks
Configure the authentication scheme	<p>Enter the default ISP domain view</p> <p>Configure the AAA scheme to be applied to the domain</p> <p>Quit to system view</p>	<p>domain <i>Domain name</i></p> <p>authentication default { hwtacacs-scheme <i>hwtacacs-scheme-name</i> [local] local none radius-scheme <i>radius-scheme-name</i> [local] }</p> <p>quit</p> <p>Optional</p> <p>By default, the local AAA scheme is applied. If you specify to apply the local AAA scheme, you need to perform the configuration concerning local user as well.</p> <p>If you specify to apply an existing scheme by providing the <i>radius-scheme-name</i> argument, you need to perform the following configuration as well:</p> <ul style="list-style-type: none"> Perform AAA-RADIUS configuration on the switch. (Refer to "AAA/RADIUS/HWTACACS Configuration" on page 747.) Configure the user name and password accordingly on the AAA server. (Refer to "Configuring RADIUS" on page 765.)
Create a local user and enter local user view	local-user <i>user-name</i>	No local user exists by default.
Set the authentication password for the local user	password { simple cipher } <i>password</i>	Required
Specify the service type for VTY users	service-type telnet [level <i>level</i>]	Required
Quit to system view	quit	-
Enter one or more VTY user interface views	user-interface vty <i>first-number</i> [<i>last-number</i>]	-
Configure to authenticate users locally or remotely	authentication-mode scheme	<p>Required</p> <p>The specified AAA scheme determines whether to authenticate users locally or remotely.</p> <p>Users are authenticated locally by default.</p>
Configure the command level available to users logging in to the user interface	user privilege level <i>level</i>	<p>Optional</p> <p>By default, commands of level 0 are available to users logging in to the VTY user interfaces.</p>
Configure the supported protocol	protocol inbound { all ssh telnet }	<p>Optional</p> <p>Both Telnet protocol and SSH protocol are supported by default.</p>
Set the command that is automatically executed when a user logs into the user interface	auto-execute command <i>text</i>	<p>Optional</p> <p>By default, no command is automatically executed when a user logs into a user interface.</p>
Define a shortcut key for aborting tasks	escape-key { default <i>character</i> }	<p>Optional</p> <p>The default shortcut key combination for aborting tasks is Ctrl+C.</p>

To do...	Use the command...	Remarks
Make terminal services available	shell	Optional Terminal services are available in all use interfaces by default.
Set the maximum number of lines the screen can contain	screen-length <i>screen-length</i>	Optional By default, the screen can contain up to 24 lines. You can use the screen-length 0 command to disable the function to display information in pages.
Set history command buffer size	history-command max-size <i>value</i>	Optional The default history command buffer size is 10. That is, a history command buffer can store up to 10 commands by default.
Set the timeout time for the user interface	idle-timeout <i>minutes</i> [<i>seconds</i>]	Optional The default timeout time of a user interface is 10 minutes. With the timeout time being 10 minutes, the connection to a user interface is terminated if no operation is performed in the user interface within 10 minutes. You can use the idle-timeout 0 command to disable the timeout function.

Note that if you configure to authenticate the users in the scheme mode, the command level available to users logging in to a switch depends on the **authentication-mode scheme [command-authorization]** command, the **user privilege level *level*** command, and the **service-type { ftp [ftp-directory *directory*] | lan-access | { ssh | telnet | terminal }* [level *level*] }** command, as listed in Table 15.

Table 15 Determine the command level when users logging in to switches are authenticated in the scheme mode

Scenario			
Authentication mode	User type	Command	Command level
Scheme (authentication -mode scheme [command-aut horization])	VTY users that are AAA-RADIUS authenticated or locally authenticated	The user privilege level level command is not executed, and the service-type command does not specify the available command level.	Level 0
		The user privilege level level command is not executed, and the service-type command specifies the available command level.	Determined by the service-type command
		The user privilege level level command is executed, and the service-type command does not specify the available command level.	Level 0
		The user privilege level level command is executed, and the service-type command specifies the available command level.	Determined by the service-type command
	VTY users that are authenticated in the RSA mode of SSH	The user privilege level level command is not executed, and the service-type command does not specify the available command level.	Level 0
		The user privilege level level command is not executed, and the service-type command specifies the available command level.	
		The user privilege level level command is executed, and the service-type command does not specify the available command level.	Determined by the user privilege level level level command
		The user privilege level level command is executed, and the service-type command specifies the available command level.	
	VTY users that are authenticated in the password mode of SSH	The user privilege level level command is not executed, and the service-type command does not specify the available command level.	Level 0
		The user privilege level level command is not executed, and the service-type command specifies the available command level.	Determined by the service-type command
		The user privilege level level command is executed, and the service-type command does not specify the available command level.	Level 0
		The user privilege level level command is executed, and the service-type command specifies the available command level.	Determined by the service-type command



Refer to “AAA/RADIUS/HWTACACS Configuration” on page 747 and “SSH Configuration” on page 1107.

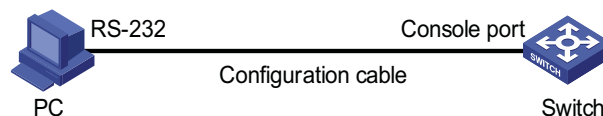
Configuration Example Network requirements

Assume that you are a level 3 AUX user and want to perform the following configuration for Telnet users logging in to VTY 0:

- Configure the name of the local user to be “guest”.
- Set the authentication password of the local user to 123456 (in plain text).
- Set the service type of VTY users to Telnet.
- Configure to authenticate users logging in to VTY 0 in scheme mode.
- The commands of level 2 are available to users logging in to VTY 0.
- Telnet protocol is supported in VTY 0.
- The screen can contain up to 30 lines.
- The history command buffer can store up to 20 commands.
- The timeout time of VTY 0 is 6 minutes.

Network diagram

Figure 12 Network diagram for Telnet configuration (with the authentication mode being scheme)

**Configuration procedure**

Enter system view, and enable the Telnet service.

```
<SW4800G> system-view
[SW4800G] telnet server enable
```

Create a local user named “guest” and enter local user view.

```
[SW4800G] local-user guest
```

Set the authentication password of the local user to 123456 (in plain text).

```
[SW4800G-luser-guest] password simple 123456
```

Set the service type to Telnet, Specify commands of level 2 are available to users logging in to VTY 0.

```
[SW4800G-luser-guest] service-type telnet level 2
```

Enter VTY 0 user interface view.

```
[SW4800G] user-interface vty 0
```

Configure to authenticate users logging in to VTY 0 in the scheme mode.

```
[SW4800G-ui-vty0] authentication-mode scheme
```

Configure Telnet protocol is supported.

```
[SW4800G-ui-vty0] protocol inbound telnet
```

```
# Set the maximum number of lines the screen can contain to 30.

[SW4800G-ui-vty0] screen-length 30

# Set the maximum number of commands the history command buffer can store
to 20.

[SW4800G-ui-vty0] history-command max-size 20

# Set the timeout time to 6 minutes.

[SW4800G-ui-vty0] idle-timeout 6
```

Telnet Connection Establishment

Telnetting to a Switch from a Terminal

You can telnet to a switch and then configure the switch if the interface of the management VLAN of the switch is assigned with an IP address. (By default, VLAN 1 is the management VLAN.)

Following are procedures to establish a Telnet connection to a switch:

Step 1: Log in to the switch through the console port, enable the Telnet server function and assign an IP address to the management VLAN interface of the switch.

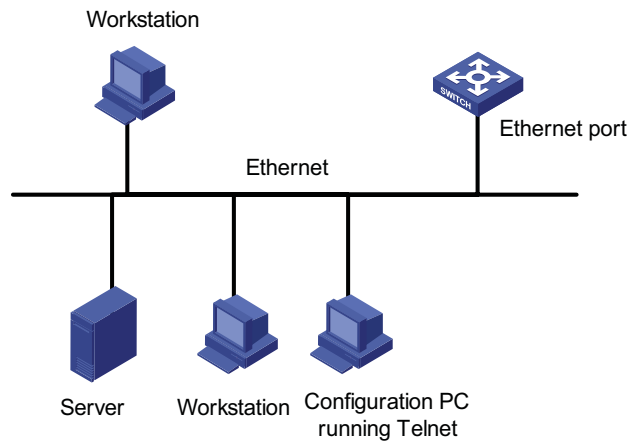
- Connect to the console port. Refer to section “Setting Up the Connection to the Console Port” on page 31.
- Execute the following commands in the terminal window to enable the Telnet server function and assign an IP address to the management VLAN interface of the switch.

```
# Enable the Telnet server function and configure the IP address of the
management VLAN interface as 202.38.160.92, and the subnet mask as
255.255.255.0.
```

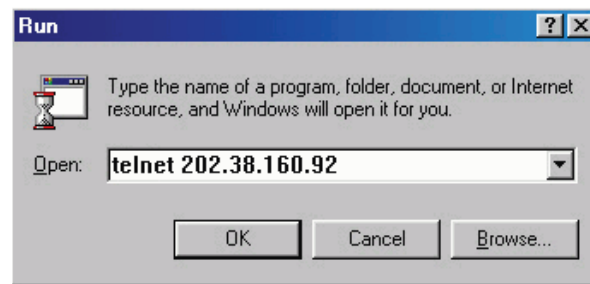
```
<SW4800G> system-view
[SW4800G] telnet server enable
[SW4800G] interface vlan-interface 1
[SW4800G-Vlan-interface1] ip address 202.38.160.92 255.255.255.0
```

Step 2: Before Telnet users can log in to the switch, corresponding configurations should have been performed on the switch according to different authentication modes for them. Refer to section “Telnet Configuration with Authentication Mode Being None” on page 49, section “Telnet Configuration with Authentication Mode Being Password” on page 52, and section “Telnet Configuration with Authentication Mode Being Scheme” on page 54 for more. By default, Telnet users need to pass the password authentication to login.

Step 3: Connect your PC to the Switch, as shown in Figure 13. Make sure the Ethernet port to which your PC is connected belongs to the management VLAN of the switch and the route between your PC and the switch is available.

Figure 13 Network diagram for Telnet connection establishment

Step 4: Launch Telnet on your PC, with the IP address of the management VLAN interface of the switch as the parameter, as shown in the following figure.

Figure 14 Launch Telnet

Step 5: Enter the password when the Telnet window displays "Login authentication" and prompts for login password. The CLI prompt (such as <SW4800G>) appears if the password is correct. If all VTY user interfaces of the switch are in use, you will fail to establish the connection and receive the message that says "All user interfaces are used, please try later!". A 3Com series Ethernet switch can accommodate up to five Telnet connections at same time.

Step 6: After successfully Telnetting to a switch, you can configure the switch or display the information about the switch by executing corresponding commands. You can also type ? at any time for help. Refer to the following chapters for the information about the commands.



- A Telnet connection will be terminated if you delete or modify the IP address of the VLAN interface in the Telnet session.
- By default, commands of level 0 are available to Telnet users authenticated by password. Refer to "Configuring User Levels and Command Levels" on page 1026 for information about command hierarchy.

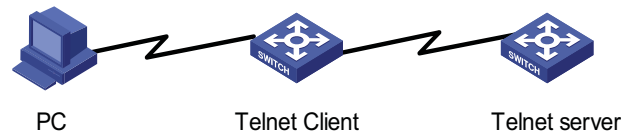
Telnetting to Another Switch from the Current Switch

You can Telnet to another switch from the current switch. In this case, the current switch operates as the client, and the other operates as the server. If the interconnected Ethernet ports of the two switches are in the same LAN segment, make sure the IP addresses of the two management VLAN interfaces to which the

two Ethernet ports belong to are of the same network segment, or the route between the two VLAN interfaces is available.

As shown in Figure 15, after Telnetting to a switch (labeled as Telnet client), you can Telnet to another switch (labeled as Telnet server) by executing the **telnet** command and then to configure the later.

Figure 15 Network diagram for Telnetting to another switch from the current switch



Step 1: Configure the user name and password for Telnet on the switch operating as the Telnet server. Refer to section “Telnet Configuration with Authentication Mode Being None” on page 49, section “Telnet Configuration with Authentication Mode Being Password” on page 52, and section “Telnet Configuration with Authentication Mode Being Scheme” on page 54 for more. By default, Telnet users need to pass the password authentication to login.

Step 2: Telnet to the switch operating as the Telnet client.

Step 3: Execute the following command on the switch operating as the Telnet client:

```
<SW4800G> telnet xxxx
```

Where xxxx is the IP address or the host name of the switch operating as the Telnet server. You can use the **ip host** to assign a host name to a switch.

Step 4: Enter the password. If the password is correct, the CLI prompt (such as <SW4800G>) appears. If all VTY user interfaces of the switch are in use, you will fail to establish the connection and receive the message that says “All user interfaces are used, please try later!”.

Step 5: After successfully Telnetting to the switch, you can configure the switch or display the information about the switch by executing corresponding commands. You can also type ? at any time for help. Refer to the following chapters for the information about the commands.

4

LOGGING IN USING MODEM

Introduction

The administrator can log in to the console port of a remote switch using a modem through PSTN (public switched telephone network) if the remote switch is connected to the PSTN through a modem to configure and maintain the switch remotely. When a network operates improperly or is inaccessible, you can log in to the switches in the network in this way to configure these switches, to query logs and warning messages, and to locate problems.

To log in to a switch in this way, you need to configure the terminal and the switch properly, as listed in the following table.

Table 16 Requirements for logging in to a switch using a modem

Item	Requirement
Administrator side	The PC can communicate with the modem connected to it. The modem is properly connected to PSTN. The telephone number of the switch side is available.
Switch side	The modem is connected to the console port of the switch properly. The modem is properly configured. The modem is properly connected to PSTN and a telephone set. The authentication mode and other related settings are configured on the switch. Refer to Table 6.

Configuration on the Administrator Side

The PC can communicate with the modem connected to it. The modem is properly connected to PSTN. And the telephone number of the switch side is available.

Configuration on the Switch Side

Modem Configuration

Perform the following configuration on the modem directly connected to the switch:

```
AT&F ----- Restore the factory settings
ATS0=1 ----- Configure to answer automatically after the first ring
AT&D ----- Ignore DTR signal
AT&K0 ----- Disable flow control
AT&R1 ----- Ignore RTS signal
AT&S0 ----- Set DSR to high level by force
ATEQ1&W ----- Disable the modem from returning command response and the result,
                save the changes
```

You can verify your configuration by executing the **AT&V** command.



The above configuration is unnecessary to the modem on the administrator side.

The configuration commands and the output of different modems may differ. Refer to the user manual of the modem when performing the above configuration.

Switch Configuration



After logging in to a switch through its console port by using a modem, you will enter the AUX user interface. The corresponding configuration on the switch is the same as those when logging in to the switch locally through its console port except that

- When you log in through the console port using a modem, the baud rate of the console port is usually set to a value lower than the transmission speed of the modem. Otherwise, packets may get lost.
- Other settings of the console port, such as the check mode, the stop bits, and the data bits, remain the default.

The configuration on the switch depends on the authentication mode the user is in. Refer to Table 6 for the information about authentication mode configuration.

Configuration on switch when the authentication mode is none

Refer to section “Console Port Login Configuration with Authentication Mode Being None” on page 35.

Configuration on switch when the authentication mode is password

Refer to section “Console Port Login Configuration with Authentication Mode Being Password” on page 38

Configuration on switch when the authentication mode is scheme

Refer to section “Console Port Login Configuration with Authentication Mode Being Scheme” on page 41.

Modem Connection Establishment

Step 1: Configure the user name and password on the switch. Refer to section “Console Port Login Configuration with Authentication Mode Being None” on page 35, section “Console Port Login Configuration with Authentication Mode Being Password” on page 38, and section “Console Port Login Configuration with Authentication Mode Being Scheme” on page 41 for more.

Step 2: Perform the following configuration on the modem directly connected to the switch.

```
AT&F ----- Restore the factory settings
ATS0=1 ----- Configure to answer automatically after the first ring
AT&D ----- Ignore DTR signal
AT&K0 ----- Disable flow control
AT&R1 ----- Ignore RTS signal
AT&S0 ----- Set DSR to high level by force
ATEQ1&W ----- Disable the modem from returning command response and the result,
                save the changes
```

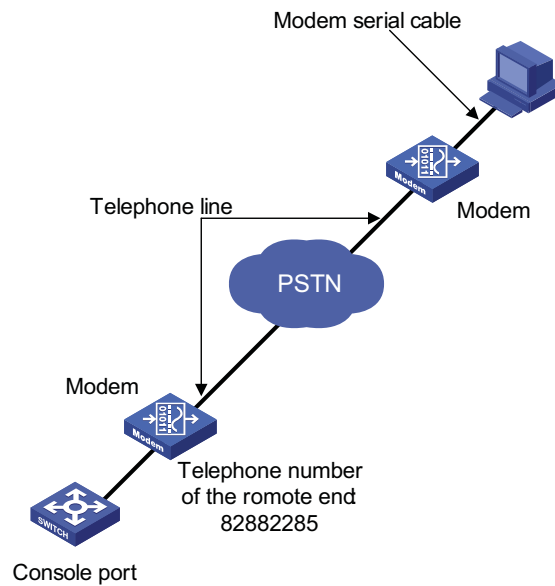
You can verify your configuration by executing the **AT&V** command.



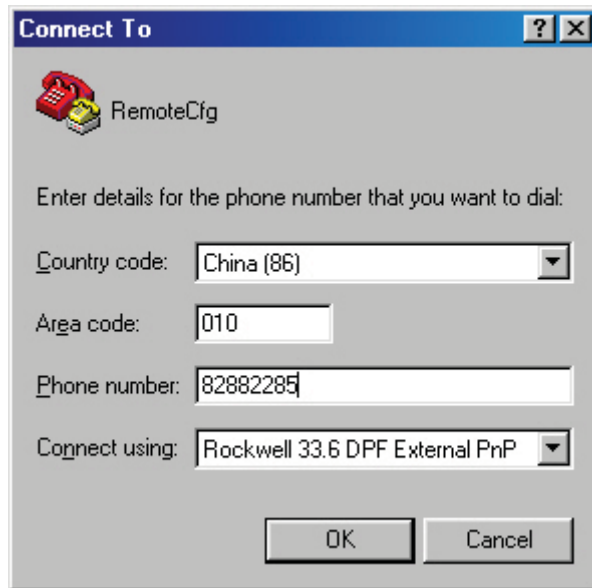
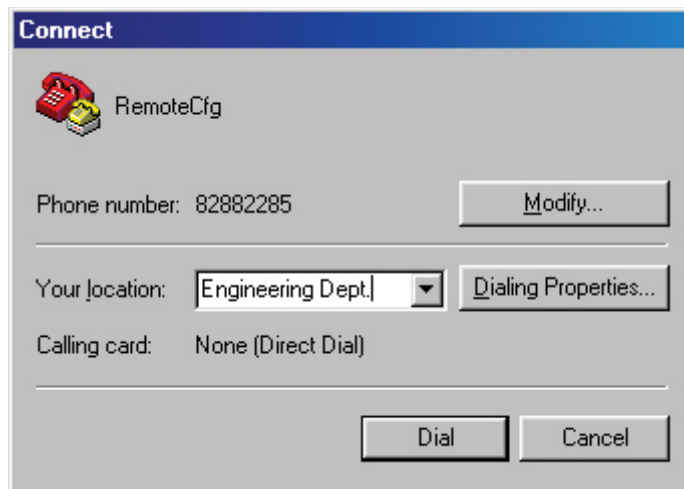
- The configuration commands and the output of different modems may differ. Refer to the user manual of the modem when performing the above configuration.
- It is recommended that the baud rate of the AUX port (also the console port) be set to a value lower than the transmission speed of the modem. Otherwise, packets may get lost.

Step 3: Connect your PC, the modems, and the switch, as shown in the following figure.

Figure 16 Establish the connection by using modems



Step 4: Launch a terminal emulation utility on the PC and set the telephone number to call the modem directly connected to the switch, as shown in Figure 17 and Figure 18. Note that you need to set the telephone number to that of the modem directly connected to the switch.

Figure 17 Set the telephone number**Figure 18** Call the modem

Step 5: Provide the password when prompted. If the password is correct, the prompt (such as <SW4800G>) appears. You can then configure or manage the switch. You can also enter the character ? at anytime for help. Refer to the following chapters for information about the configuration commands.



If you perform no AUX user-related configuration on the switch, the commands of level 3 are available to modem users. Refer to “Configuring User Levels and Command Levels” on page 1026 for information about command level.

5

LOGGING IN THROUGH WEB-BASED NETWORK MANAGEMENT SYSTEM

Introduction

A Switch 4800G has a Web server built in. You can log in to a Switch 4800G through a Web browser and manage and maintain the switch intuitively by interacting with the built-in Web server.

To log in to a Switch 4800G through the built-in Web-based network management system, you need to perform the related configuration on both the switch and the PC operating as the network management terminal.

Table 17 Requirements for logging in to a switch through the Web-based network management system

Item	Requirement
Switch	Start the Web server The IP address of the management VLAN of the switch is configured. The route between the switch and the network management terminal is available. (Refer to "IP Addressing Configuration" on page 121 and "IP Routing Overview" on page 241 for more.) The user name and password for logging in to the Web-based network management system are configured.
PC operating as the network management terminal	IE is available. The IP address of the management VLAN interface of the switch is available.

HTTP Connection Establishment

Step 1: Log in to the switch through the console port and assign an IP address to the management VLAN interface of the switch. By default, VLAN 1 is the management VLAN.

- Connect to the console port. Refer to section "Setting Up the Connection to the Console Port" on page 31.
- Execute the following commands in the terminal window to assign an IP address to the management VLAN interface of the switch.

Configure the IP address of the management VLAN interface to be 10.153.17.82 with the mask 255.255.255.0.

```
<SW4800G> system-view
[SW4800G] interface vlan-interface 1
[SW4800G-Vlan-interface1] ip address 10.153.17.82 255.255.255.0
```

Step 2: Configure the user name and the password for the Web-based network management system.

```
# Configure the user name to be admin.

[SW4800G] local-user admin

# Set the user level to level 3.

[SW4800G-luser-admin] service-type telnet level 3

# Set the password to admin.

[SW4800G-luser-admin] password simple admin
```

Step 3: Establish an HTTP connection between your PC and the switch, as shown in the following figure.

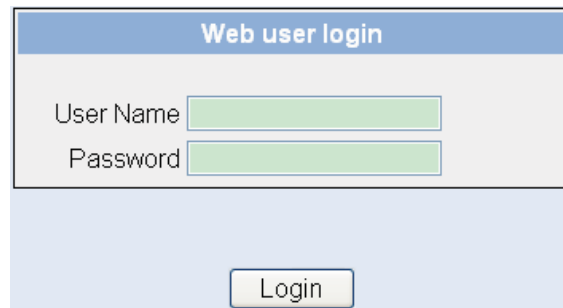
Figure 19 Establish an HTTP connection between your PC and the switch



Step 4: Log in to the switch through IE. Launch IE on the Web-based network management terminal (your PC) and enter the IP address of the management VLAN interface of the switch (here it is http://10.153.17.82). (Make sure the route between the Web-based network management terminal and the switch is available.)

Step 5: When the login interface (shown in Figure 20) appears, enter the user name and the password configured in step 2 and click <Login> to bring up the main page of the Web-based network management system.

Figure 20 The login page of the Web-based network management system



Web Server Shutdown/Startup

You can shut down or start up the Web server.

To do...	Use the command...	Remarks
Enter system view	system-view	-
Shut down the Web server	undo ip http enable	Required Execute this command in system view. The Web server is started by default.

To do...	Use the command...	Remarks
Start the Web server	ip http enable	Required Execute this command in system view.

Displaying Web Users

After the above configurations, execute the **display** command in any view to display the information about Web users, and thus to verify the configuration effect.

Table 18 Display information about Web users

To do...	Use the command...
Display information about Web users	display web users

6

LOGGING IN THROUGH NMS

Introduction

You can also log in to a switch through an NMS (network management station), and then configure and manage the switch through the agent module on the switch.

- The agent here refers to the software running on network devices (switches) and as the server.
- SNMP (simple network management protocol) is applied between the NMS and the agent.

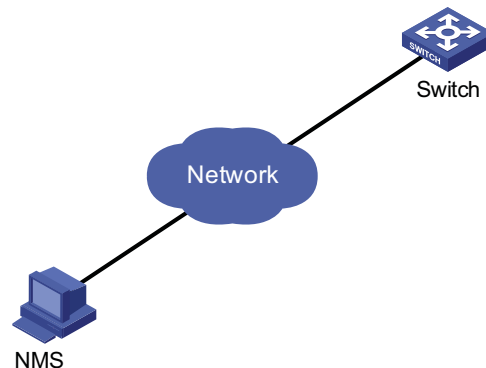
To log in to a switch through an NMS, you need to perform related configuration on both the NMS and the switch.

Table 19 Requirements for logging in to a switch through an NMS

Item	Requirement
Switch	The IP address of the management VLAN of the switch is configured. The route between the NMS and the switch is available. (Refer to the module "IP Addressing Configuration" on page 121 and "IP Routing Overview" on page 241 for more.) The basic SNMP functions are configured. (Refer to the "SNMP Configuration" on page 931 for more.)
NMS	The NMS is properly configured. (Refer to the user manual of your NMS for more.)

Connection Establishment Using NMS

Figure 21 Network diagram for logging in through an NMS



7

CONFIGURING SOURCE IP ADDRESS FOR TELNET SERVICE PACKETS

Go to these sections for information you are interested in:

- “Overview” on page 73
- “Configuring Source IP Address for Telnet Service Packets” on page 73
- “Displaying the source IP address/Interface Specified for Telnet Packets” on page 74

Overview

You can configure source IP address or source interface for Telnet client. This provides a way to manage services and enhances security.

The source IP address specified for Telnet service packets is the IP address of an Loopback interface or VLAN interface. After you specify the IP address of a virtual Loopback interface or an unused VLAN interface as the source IP address of Telnet service packets, the IP address is used as the source IP address no matter which interface of the switch is used to transmit packets between the Telnet client and the Telnet server. This conceals the IP address of the actual interface used. As a result, external attacks are guarded and the security is improved. On the other hand, you can configure the Telnet server to accept only Telnet service packets with specific source IP addresses to make sure specific users can log in to the switch.

Configuring Source IP Address for Telnet Service Packets

This feature can be configured in either user view or system view. The configuration performed in user view takes effect for only the current session, while the configuration performed in system view takes effect for all the following sessions.

Configuration in user view

Table 20 Configure a source IP address for service packets in user view

To do...	Use the command...	Remarks
Specify the source IP address or source interface for the switch for it to log in to another device as a Telnet client	telnet <i>remote-system</i> [<i>port-number</i>] [source { ip <i>ip-address</i> interface <i>interface-type</i> <i>interface-number</i> }]	Optional Not specified by default

Configuration in system view

Table 21 Configure a source IP address for service packets in system view

To do...	Use the command...	Remarks
Enter system view	system-view	-

Table 21 Configure a source IP address for service packets in system view

To do...	Use the command...	Remarks
Specify the source IP address or source interface for the switch for it to log in to another device as a Telnet client	telnet client source { ip <i>ip-address</i> interface <i>interface-type interface-number</i> }	Optional Not specified by default



To perform the configurations listed in Table 20 and Table 21, make sure that

- The IP address specified is that of the local device.
- The interface specified exists.
- If a source IP address (or source interface) is specified, you need to make sure that the route between the IP addresses (or interface) of both sides is reachable.

Displaying the source IP address/Interface Specified for Telnet Packets

Follow these steps to display the source IP address/interface specified for Telnet packets:

To do...	Use the command...
Display the source IP address/interface specified for Telnet packets	display telnet client configuration

8

CONTROLLING LOGIN USERS

Introduction

A switch provides ways to control different types of login users, as listed in Table 22.

Table 22 Ways to control different types of login users

Login mode	Control method	Implementation	Related section
Telnet	By source IP addresses	Through basic ACLs	Section "Controlling Telnet Users by Source IP Addresses" on page 75.
	By source and destination IP addresses	Through advanced ACLs	Section "Controlling Telnet Users by Source and Destination IP Addresses" on page 76.
	By source MAC addresses	Through Layer 2 ACLs	Section "Controlling Telnet Users by Source MAC Addresses" on page 76
SNMP	By source IP addresses	Through basic ACLs	Section "Controlling Network Management Users by Source IP Addresses" on page 78
WEB	By source IP addresses	Through basic ACLs	Section "Controlling Web Users by Source IP Addresses" on page 80
	Disconnect Web users by force	By executing commands in CLI	Section "Disconnecting a Web User by Force" on page 80.

Controlling Telnet Users

Prerequisites The controlling policy against Telnet users is determined, including the source and destination IP addresses to be controlled and the controlling actions (permitting or denying).

Controlling Telnet Users by Source IP Addresses Controlling Telnet users by source IP addresses is achieved by applying basic ACLs, which are numbered from 2000 to 2999.

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a basic ACL or enter basic ACL view	acl [ipv6] number acl-number [match-order { config auto }]	As for the acl number command, the config keyword is specified by default.

To do...	Use the command...	Remarks
Define rules for the ACL	rule [<i>rule-id</i>] { permit deny } [source { <i>sour-addr</i> <i>sour-wildcard</i> any }] [time-range <i>time-name</i>] [fragment logging]*	Required
Quit to system view	quit	-
Enter user interface view	user-interface [<i>type</i>] [<i>first-number</i> [<i>last-number</i>]]	-
Apply the ACL to control Telnet users by source IP addresses	acl [ipv6] <i>acl-number</i> { inbound outbound }	Required The inbound keyword specifies to filter the users trying to Telnet to the current switch. The outbound keyword specifies to filter users trying to Telnet to other switches from the current switch.

Controlling Telnet Users by Source and Destination IP Addresses

Controlling Telnet users by source and destination IP addresses is achieved by applying advanced ACLs, which are numbered from 3000 to 3999. Refer to “ACL Overview” on page 835 for information about defining an ACL.

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create an advanced ACL or enter advanced ACL view	acl [ipv6] number <i>acl-number</i> [match-order { config auto }]	As for the acl number command, the config keyword is specified by default.
Define rules for the ACL	rule [<i>rule-id</i>] { permit deny } <i>rule-string</i>	Required You can define rules as needed to filter by specific source and destination IP addresses.
Quit to system view	quit	-
Enter user interface view	user-interface [<i>type</i>] [<i>first-number</i> [<i>last-number</i>]]	-
Apply the ACL to control Telnet users by specified source and destination IP addresses	acl [ipv6] <i>acl-number</i> { inbound outbound }	Required The inbound keyword specifies to filter the users trying to Telnet to the current switch. The outbound keyword specifies to filter users trying to Telnet to other switches from the current switch.

Controlling Telnet Users by Source MAC Addresses

Controlling Telnet users by source MAC addresses is achieved by applying Layer 2 ACLs, which are numbered from 4000 to 4999. Refer to “Configuring an Ethernet Frame Header ACL” on page 845 for information about defining an ACL.

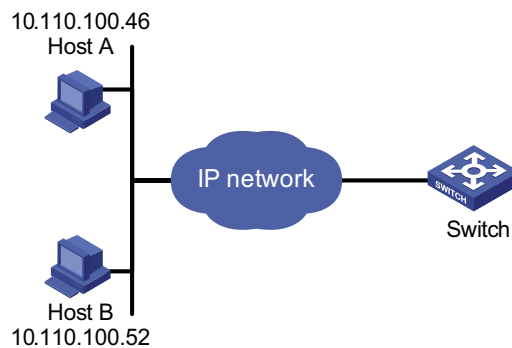
To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a basic ACL or enter basic ACL view	acl number <i>acl-number</i> [match-order { config auto }]	As for the acl number command, the config keyword is specified by default.
Define rules for the ACL	rule [<i>rule-id</i>] { permit deny } <i>rule-string</i>	Required You can define rules as needed to filter by specific source MAC addresses.
Quit to system view	quit	-
Enter user interface view	user-interface [<i>type</i>] <i>first-number</i> [<i>last-number</i>]	-
Apply the ACL to control Telnet users by source MAC addresses	acl <i>acl-number</i> inbound	Required The inbound keyword specifies to filter the users trying to Telnet to the current switch.

Configuration Example Network requirements

Only the Telnet users sourced from the IP address of 10.110.100.52 and 10.110.100.46 are permitted to log in to the switch.

Network diagram

Figure 22 Network diagram for controlling Telnet users using ACLs



Configuration procedure

Define a basic ACL.

```
<SW4800G> system-view
[SW4800G] acl number 2000 match-order config
[SW4800G-acl-basic-2000] rule 1 permit source 10.110.100.52 0
[SW4800G-acl-basic-2000] rule 2 permit source 10.110.100.46 0
[SW4800G-acl-basic-2000] rule 3 deny source any
[SW4800G-acl-basic-2000] quit
```

Apply the ACL.

```
[SW4800G] user-interface vty 0 4
[SW4800G-ui-vty0-4] acl 2000 inbound
```

Controlling Network Management Users by Source IP Addresses

You can manage a Switch 4800G through network management software. Network management users can access switches through SNMP.

You need to perform the following two operations to control network management users by source IP addresses.

- Defining an ACL
- Applying the ACL to control users accessing the switch through SNMP

Prerequisites

The controlling policy against network management users is determined, including the source IP addresses to be controlled and the controlling actions (permitting or denying).

Controlling Network Management Users by Source IP Addresses

Controlling network management users by source IP addresses is achieved by applying basic ACLs, which are numbered from 2000 to 2999. Refer to “Configuring an Advanced IPv4 ACL” on page 844 for information about defining an ACL.

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a basic ACL or enter basic ACL view	acl number <i>acl-number</i> [match-order { config auto }]	As for the acl number command, the config keyword is specified by default.
Define rules for the ACL	rule [<i>rule-id</i>] { permit deny } [source { <i>sour-addr</i> <i>sour-wildcard</i> any }] [time-range <i>time-name</i> fragment logging]*	Required
Quit to system view	quit	-
Apply the ACL while configuring the SNMP community name	snmp-agent community { read write } <i>community-name</i> [mib-view <i>view-name</i> acl <i>acl-number</i>]*	Required
Apply the ACL while configuring the SNMP group name	snmp-agent group { v1 v2c } <i>group-name</i> [read-view <i>read-view</i>] [write-view <i>write-view</i>] [notify-view <i>notify-view</i>] [acl <i>acl-number</i>] snmp-agent group v3 <i>group-name</i> [authentication privacy] [read-view <i>read-view</i>] [write-view <i>write-view</i>] [notify-view <i>notify-view</i>] [acl <i>acl-number</i>]	Required
Apply the ACL while configuring the SNMP user name	snmp-agent usm-user { v1 v2c } <i>user-name</i> <i>group-name</i> [acl <i>acl-number</i>] snmp-agent usm-user v3 <i>user-name</i> <i>group-name</i> [authentication-mode { md5 sha } <i>auth-password</i> [privacy-mode { des56 aes128 } <i>priv-password</i>]] [acl <i>acl-number</i>]	Required



You can specify different ACLs while configuring the SNMP community name, the SNMP group name and the SNMP user name.

As SNMP community name is a feature of SNMPv1 and SNMPv2c, the specified ACLs in the command that configures SNMP community names (the **snmp-agent community** command) take effect in the network management systems that adopt SNMPv1 or SNMPv2c.

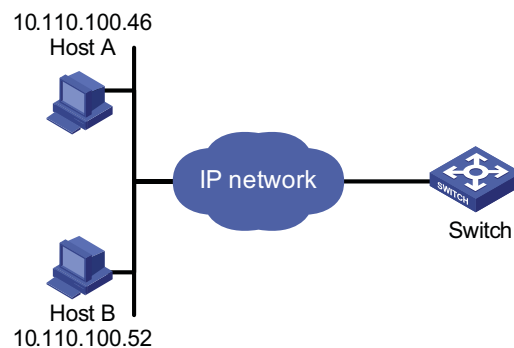
Similarly, as SNMP group name and SNMP user name are features of SNMPv2c and the higher SNMP versions, the specified ACLs in the commands that configure SNMP group names (the **snmp-agent group** command and the **snmp-agent group v3** command) and SNMP user names (the **snmp-agent usm-user** command and the **snmp-agent usm-user v3** command) take effect in the network management systems that adopt SNMPv2c or higher SNMP versions. If you configure both the SNMP group name and the SNMP user name and specify ACLs in the two operations, the switch will filter network management users by both SNMP group name and SNMP user name.

Configuration Example Network requirements

Only SNMP users sourced from the IP addresses of 10.110.100.52 and 10.110.100.46 are permitted to access the switch.

Network diagram

Figure 23 Network diagram for controlling SNMP users using ACLs



Configuration procedure

Define a basic ACL.

```
<SW4800G> system-view
[SW4800G] acl number 2000 match-order config
[SW4800G-acl-basic-2000] rule 1 permit source 10.110.100.52 0
[SW4800G-acl-basic-2000] rule 2 permit source 10.110.100.46 0
[SW4800G-acl-basic-2000] rule 3 deny source any
[SW4800G-acl-basic-2000] quit
```

Apply the ACL to only permit SNMP users sourced from the IP addresses of 10.110.100.52 and 10.110.100.46 to access the switch.

```
[SW4800G] snmp-agent community read h3c acl 2000
[SW4800G] snmp-agent group v2c h3cgroup acl 2000
[SW4800G] snmp-agent usm-user v2c h3cuser h3cgroup acl 2000
```

Controlling Web Users by Source IP Address

You can manage a Switch 4800G remotely through Web. Web users can access a switch through HTTP connections.

You need to perform the following two operations to control Web users by source IP addresses.

- Defining an ACL
- Applying the ACL to control Web users

Prerequisites The controlling policy against Web users is determined, including the source IP addresses to be controlled and the controlling actions (permitting or denying).

Controlling Web Users by Source IP Addresses Controlling Web users by source IP addresses is achieved by applying basic ACLs, which are numbered from 2000 to 2999.

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a basic ACL or enter basic ACL view	acl number <i>acl-number</i> [match-order { config auto }]	As for the acl number command, the config keyword is specified by default.
Define rules for the ACL	rule [<i>rule-id</i>] { permit deny } [source { <i>sour-addr</i> <i>sour-wildcard</i> any } time-range <i>time-name</i> fragment logging]*	Required
Quit to system view	quit	-
Apply the ACL to control Web users	ip http acl <i>acl-number</i>	Optional

Disconnecting a Web User by Force

The administrator can disconnect a Web user by force using the related command.

To do...	Use the command...	Remarks
Disconnect a Web user by force	free web-users { all user-id <i>user-id</i> user-name <i>user-name</i> }	Required Execute this command in user view.

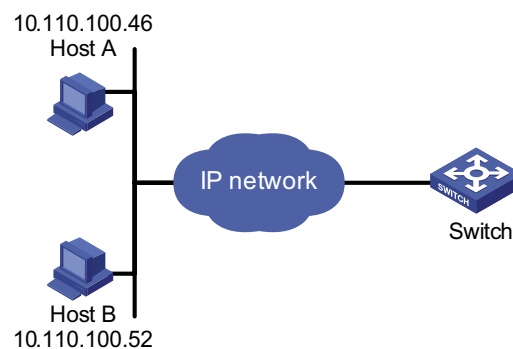
Configuration Example

Network requirements

Only the users sourced from the IP address of 10.110.100.52 are permitted to access the switch.

Network diagram

Figure 24 Network diagram for controlling Web users using ACLs



Configuration procedure

Define a basic ACL.

```
<SW4800G> system-view  
[SW4800G] acl number 2030 match-order config  
[SW4800G-acl-basic-2030] rule 1 permit source 10.110.100.52 0  
[SW4800G-acl-basic-2030] rule 2 deny source any
```

Apply the ACL to only permit the Web users sourced from the IP address of 10.110.100.52 to access the switch.

```
[SW4800G] ip http acl 2030
```


9

VLAN CONFIGURATION

When configuring VLAN, go to these sections for information you are interested in:

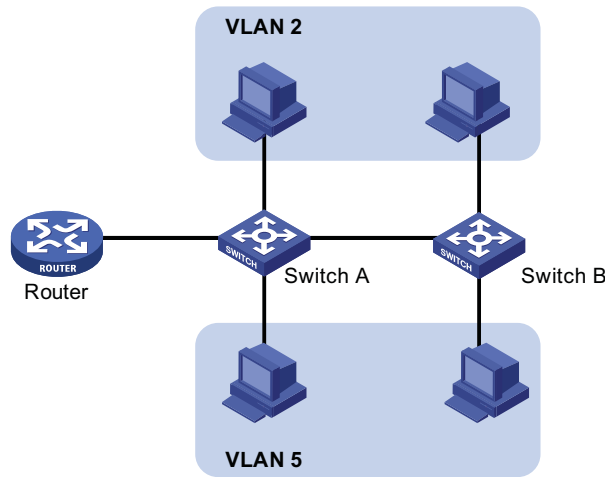
- "Introduction to VLAN" on page 83
- "Configuring Basic VLAN Attributes" on page 86
- "Basic VLAN Interface Configuration" on page 86
- "Port-Based VLAN Configuration" on page 87
- "MAC Address-Based VLAN Configuration" on page 91
- "Protocol-Based VLAN Configuration" on page 92
- "Configuring IP-Subnet-Based VLAN" on page 94
- "Displaying and Maintaining VLAN" on page 95
- "VLAN Configuration Example" on page 95

Introduction to VLAN

VLAN Overview

Ethernet is a network technology based on the Carrier Sense Multiple Access/Collision Detect (CSMA/CD) mechanism. As the medium is shared in an Ethernet, network performance may degrade as the number of hosts on the network is increasing. If the number of the hosts in the network reaches a certain level, problems caused by collisions, broadcasts, and so on emerge, which may cause the network operating improperly. In addition to the function that suppresses collisions (which can also be achieved by interconnecting LANs), virtual LAN (VLAN) can also isolate broadcast packets. VLAN divides a LAN into multiple logical LANs with each being a broadcast domain. Hosts in the same VLAN can communicate with each other like in a LAN. However, hosts from different VLANs cannot communicate directly. In this way, broadcast packets are confined to a single VLAN, as illustrated in the following figure.

Figure 25 A VLAN diagram



A VLAN is not restricted by physical factors, that is to say, hosts that reside in different network segments may belong to the same VLAN, users in a VLAN can be connected to the same switch, or span across multiple switches or routers.

VLAN technology has the following advantages:

- 1 Broadcast traffic is confined to each VLAN, reducing bandwidth utilization and improving network performance.
- 2 LAN security is improved. Packets in different VLANs are isolated at Layer 2. That is, users in a VLAN cannot communicate with users in other VLANs directly, unless routers or Layer 3 switches are used.
- 3 A more flexible way to establish virtual workgroups. With VLAN technology, a virtual workgroup can be created spanning physical network segments. That is, users from the same workgroup do not have to be within the same physical area, making network construction and maintenance much easier and more flexible.

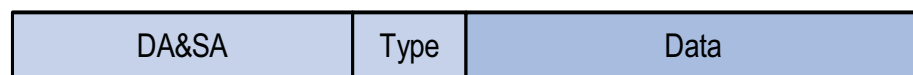
VLAN Fundamental

To enable packets being distinguished by the VLANs they belong to, The VLAN tag fields used to identify VLANs are added to packets. As common switches operate on the data link layer of the OSI model, they only process data link layer encapsulation information and the VLAN tag thus needs to be inserted to the data link layer encapsulation.

The format of the packets carrying the VLAN tag fields is defined in IEEE 802.1Q, which is issued by IEEE in 1999.

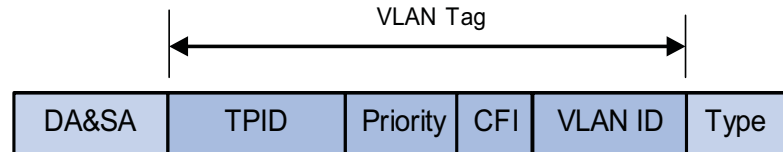
In the header of a traditional Ethernet data frame, the field following the destination MAC address and the source MAC address is the Type field, which indicates the upper layer protocol type. Figure 26 illustrates the format of a traditional Ethernet frame, where DA stands for destination MAC address, SA stands for source MAC address, and Type stands for the upper layer protocol type of the frame.

Figure 26 The format of a traditional Ethernet frame



IEEE802.1Q defines a four-byte VLAN Tag between the DA&SA field and the Type field to carry VLAN-related information, as shown in Figure 27.

Figure 27 The position and the format of the VLAN Tag



The VLAN Tag comprises four fields: the tag protocol identifier (TPID) field, the Priority field, the canonical format indicator (CFI) field, and the VLAN ID field.

- The TPID field, 16 bits in length and with a value of 0x8100, indicates that a packet carries a VLAN tag with it.
- The Priority field, three bits in length, indicates the 802.1p priority of a packet. For information about packet priority, refer to “Introduction to QoS Policies” on page 870.
- The CFI field, one bit in length, specifies whether or not the MAC addresses are encapsulated in standard format when packets are transmitted across different medium. With the field set to 0, MAC addresses are encapsulated in standard format; with the field set to 1, MAC addresses are encapsulated in non-standard format. The field is 0 by default.
- The VLAN ID field, 12 bits in length and with its value ranging from 0 to 4095, identifies the ID of the VLAN a packet belongs to. As VLAN IDs of 0 and 4095 are reserved by the protocol, the value of this field actually ranges from 1 to 4094.

A network device determines the VLAN to which a packet belongs to by the VLAN ID field the packet carries. The VLAN Tag determines the way a packet is processed. For more information, refer to section “Introduction to Port-Based VLAN” on page 87.



The frame format mentioned here is that of Ethernet II. Besides Ethernet II encapsulation, other types of encapsulation, including 802.2 LLC, 802.2 SNAP, and 802.3 raw are also supported by Ethernet. The VLAN tag fields are also added to packets adopting these encapsulation formats for VLAN identification.

VLAN Classification

Based on how VLANs are established, VLANs fall into different categories. The following types are the most commonly used:

- Port-based
- MAC address-based
- Protocol-based
- IP-subnet-based
- Policy-based
- Other types

The the Switch 4800G support port-based VLAN, MAC address-based VLAN, protocol-based VLAN, and IP-subnet-based VLAN.

Configuring Basic VLAN Attributes

Follow these steps to configure basic VLAN attributes:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create VLANs	vlan { <i>vlan-id1</i> [to <i>vlan-id2</i>] all }	Optional Using this command can create multiple VLANs in a bulk.
Enter VLAN view	vlan <i>vlan-id</i>	Required If the specified VLAN does not exist, the command creates the VLAN and then enters its view. By default, only the default VLAN (that is, VLAN 1) exists in the system.
Specify a descriptive string for the VLAN	description <i>text</i>	Optional VLAN ID used by default, for example, "VLAN 0001"



- *As the default VLAN, VLAN 1 cannot be created or removed.*
- *You cannot manually create or remove reserved VLANs, which are reserved for specific functions.*
- *Dynamic VLANs cannot be removed using the **undo vlan** command.*
- *If a VLAN has a QoS policy configured, the VLAN cannot be removed.*
- *If a VLAN is configured as a remote-probe VLAN for remote port mirroring, it cannot be removed using the **undo vlan** command unless its remote-probe VLAN configuration is removed.*

Basic VLAN Interface Configuration

Hosts of different VLANs cannot communicate directly. That is, routers or Layer 3 switches are needed for packets to travel across different VLANs. VLAN interfaces are used to forward VLAN packets on Layer 3.

VLAN interfaces are Layer 3 virtual interfaces (which do not exist physically on devices) used for Layer 3 interoperability between different VLANs. Each VLAN can have one VLAN interface. Packets of a VLAN can be forwarded on network layer through the corresponding VLAN interface. As each VLAN forms a broadcast domain, a VLAN can be an IP network segment and the VLAN interface can be the gateway to enable IP address-based Layer 3 forwarding.

Follow these steps to configure VLAN interface basic attributes:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a VLAN interface or enter VLAN interface view	interface Vlan-interface <i>vlan-interface-id</i>	Required This command leads you to VLAN interface view if the VLAN interface already exists.

To do...	Use the command...	Remarks
Configure an IP address for the VLAN interface	ip address <i>ip-address</i> { <i>mask</i> <i>mask-length</i> } [sub]	Optional Not configured by default
Specify the descriptive string for the VLAN interface	description <i>text</i>	Optional VLAN interface name is used by default, for example, "Vlan-interface1 Interface".
Bring up the VLAN interface	undo shutdown	Optional By default, a VLAN interface is up. The state of a VLAN interface also depends on the states of the ports in the VLAN. If all the ports in the VLAN are down, the VLAN interface is down; if one or more ports in the VLAN are up, the VLAN interface is up. If a VLAN interface is manually shut down, the VLAN interface is always down regardless of the states of ports in the VLAN.



Before creating a VLAN interface, ensure that the corresponding VLAN already exists. Otherwise, the specified VLAN interface will not be created.

Port-Based VLAN Configuration

Introduction to Port-Based VLAN

This is the simplest and yet the most effective way of classifying VLANs. It groups VLAN members by port. After added to a VLAN, a port can forward the packets of the VLAN.

Port link type

Based on the tag handling mode, a port's link type can be one of the following three:

- Access port: the port only belongs to one VLAN, normally used to connect user device;
- Trunk port: the port can belong to multiple VLANs, can receive/send packets for multiple VLANs, normally used to connect network devices;
- Hybrid port: the port can belong to multiple VLANs, can receive or send packets for multiple VLANs, used to connect either user or network devices;

The differences between Hybrid and Trunk port:

- A Hybrid port allows packets of multiple VLANs to be sent without the Tag label;
- A Trunk port only allows packets from the default VLAN to be sent without the Tag label.

Default VLAN

You can configure the default VLAN for a port. By default, VLAN 1 is the default VLAN for all ports. However, this can be changed as needed.

- An Access port only belongs to one VLAN. Therefore, its default VLAN is the VLAN it resides in and cannot be configured.
- You can configure the default VLAN for the Trunk port or the Hybrid port as they can both belong to multiple VLANs.
- After deletion of the default VLAN using the **undo vlan** command, the default VLAN for an Access port will revert to VLAN 1, whereas that for the Trunk or Hybrid port remains, meaning the port can use a nonexistent VLAN as the default VLAN.



For a port in automatic voice VLAN mode, do not set the voice VLAN as the default VLAN of the port. Otherwise, the system prompts error information. For information about voice VLAN, refer to “Voice VLAN Configuration” on page 99.

Configured with the default VLAN, a port handles packets in the following ways:

Inbound packets handling			
Port type	If no tag is carried in the packet	If a tag is carried in the packet	Outbound packets handling
Access Port	Tag the packet with the default VLAN ID	<ul style="list-style-type: none"> ■ Receive the packet if its VLAN ID is the same as the default VLAN ID ■ Discard the packet if its VLAN ID is different from the default VLAN ID 	Strip the Tag and send the packet as the VLAN ID is the same with the default VLAN ID
Trunk port	Check whether the default VLAN ID of the port is in the list of VLANs allowed to pass through the port, if yes, tag the packet with the default VLAN ID; if no, discard the packet	<ul style="list-style-type: none"> ■ Receive the packet if the VLAN ID is in the list of VLANs allowed to pass through the port ■ Discard the packet if the VLAN ID is not in the list of VLANs allowed to pass through the port 	<ul style="list-style-type: none"> ■ Strip the tag and send the packet if the VLAN ID is the same as the default VLAN ID ■ Keep the tag and send the packet if the VLAN ID is not the same as the default VLAN ID but allowed to pass through the port
Hybrid port			Send the packet if the VLAN ID is allowed to pass through the port. Use the port hybrid vlan command to configure whether the port keeps or strips the tags when sending packets of a VLAN (including the default VLAN).

Configuring an Access-Port-Based VLAN

There are two ways to configure Access-port-based VLAN: one way is to configure in VLAN view, the other way is to configure in Ethernet port view/port group view.

Follow these steps to configure the Access-port-based VLAN in VLAN view:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter VLAN view	vlan <i>vlan-id</i>	Required If the specified VLAN does not exist, this command be created first creates the VLAN before entering its view.
Add an Access port to the current VLAN	port <i>interface-list</i>	Required By default, system will add all ports to VLAN 1.

Follow these steps to configure the Access-port-based VLAN in Ethernet port view/port group view:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view or port group view	interface <i>interface-type interface-number</i> port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Use either command In Ethernet port view, the subsequent configurations only apply to the current port; In port group view, the subsequent configurations apply to all ports in the port group.
Configure the port link type as Access	port link-type access	Optional The link type of a port is Access by default.
Add the current Access port to a specified VLAN	port access vlan <i>vlan-id</i>	Optional By default, all Access ports belong to VLAN 1.



To add an Access port to a VLAN, make sure the VLAN already exists.

Configuring a Trunk-Port-Based VLAN

A Trunk port may belong to multiple VLANs, and you can only perform this configuration in Ethernet port view or port group view.

Follow these steps to configure the Trunk-port-based VLAN:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view or port group view	interface <i>interface-type interface-number</i> port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Use either command In Ethernet port view, the subsequent configurations only apply to the current port; in port group view, the subsequent configurations apply to all ports in the port group.
Configure the port link type as Trunk	port link-type trunk	Required

To do...	Use the command...	Remarks
Allow the specified VLANs to pass through the current Trunk port	port trunk permit vlan { <i>vlan-id-list</i> all }	Required By default, all Trunk ports only allow packets of VLAN 1 to pass.
Configure the default VLAN for the Trunk port	port trunk pvid vlan <i>vlan-id</i>	Optional VLAN 1 is the default by default.



- *To convert a Trunk port into a Hybrid port (or vice versa), you need to use the Access port as a medium. For example, the Trunk port has to be configured as an Access port first and then a Hybrid port.*
- *The default VLAN IDs of the Trunk ports on the local and peer devices must be the same. Otherwise, packets cannot be transmitted properly.*

Configuring a Hybrid-Port-Based VLAN

A Hybrid port may belong to multiple VLANs, and this configuration can only be performed in Ethernet port view or port group view.

Follow these steps to configure the Hybrid-port-based VLAN:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view or port group view	Enter Ethernet port view interface <i>interface-type interface-number</i> Enter port group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Use either command; In Ethernet port view, the subsequent configurations only apply to the current port; in port group view, the subsequent configurations apply to all ports in the port group
Configure the port link type as Hybrid	port link-type hybrid	Required
Allow the specified VLANs to pass through the current Hybrid port	port hybrid vlan <i>vlan-id-list</i> { tagged untagged }	Required By default, all Hybrid ports only allow packets of VLAN 1 to pass.
Configure the default VLAN of the Hybrid port	port hybrid pvid vlan <i>vlan-id</i>	Optional VLAN 1 is the default by default



- *To configure a Trunk port into a Hybrid port (or vice versa), you need to use the Access port as a medium. For example, the Trunk port has to be configured as an Access port first and then a Hybrid port.*
- *Ensure that the VLANs already exist before configuring them to pass through a Hybrid port.*
- *The default VLAN IDs of the Hybrid ports on the local and the peer devices must be the same. Otherwise, packets cannot be transmitted properly.*

MAC Address-Based VLAN Configuration

Introduction to MAC Address-Based VLAN

With MAC address-based VLANs created, the VLAN to which a packet belongs is determined by its source MAC address, and packets in a MAC address-based VLAN are forwarded after being tagged with the tag of the VLAN. This function is usually coupled with the security technologies (such as 802.1X) to provide secure and flexible network accesses for terminal devices.

MAC address-based VLAN implementation

With MAC address-based VLANs created on a port, the port operates as follows:

- If an untagged packet is received, the port checks its MAC address VLAN entries for the one that matches the source MAC address of the packet. If the entry exists, the packet is forwarded based on the matched VLAN ID and the precedence value; otherwise, the packet is forwarded based on other match rules.
- If a tagged packet is received, the port processes the packet in the same way as it processes port-based VLAN packets, that is, forwards the packet if the VLAN corresponding to the VLAN tag is permitted by the port or drops the packet if the VLAN corresponding to the VLAN tag is not permitted by the port.

The ways to create MAC address-based VLANs

A MAC address-based VLAN can be created in one of the following two ways.

- Static configuration (through CLI)

You can associate MAC addresses and VLANs by using corresponding commands.

- Auto configuration through the authentication server (that is, VLAN issuing)

The device associates MAC addresses and VLANs dynamically based on the information provided by the authentication server. If a user goes offline, the corresponding MAC address-to-VLAN association is removed automatically. Auto configuration requires MAC address-to-VLAN mapping relationship be configured on the authentication server. For detailed information, refer to “VLAN Assigning” on page 740.

The two configuration methods can be used at the same time, that is, you can configure a MAC address-to-VLAN entry on both the local device and the authentication server at the same time. Note that the MAC address-to-VLAN entry configuration takes effect only when the configuration on the local device is consistent with that on the authentication server.

Configuring a MAC Address-Based VLAN



MAC address-based VLANs are available only on Hybrid ports.

Follow these steps to configure a MAC address-based VLAN:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Associate MAC addresses with a VLAN	mac-vlan mac-address <i>mac-addr</i> [mask mac-mask] vlan vlan-id [priority priority]	Required
Enter Ethernet interface view or port group view	Enter Ethernet interface view interface interface-type interface-number Enter port group view port-group { manual port-group-name aggregation agg-id }	Use either command. The configuration performed in Ethernet interface view applies to the current port only; the configuration performed in port group view applies to all the ports in the port group.
Configure the link type of the port(s) as hybrid	port link-type hybrid	Required
Configure the current hybrid port(s) to permit packets of specific MAC address-based VLANs	port hybrid vlan vlan-id-list { tagged untagged }	Required By default, a hybrid port only permits the packets of VLAN 1.
Enable MAC address-based VLAN	mac-vlan enable	Required Disabled by default
Configure VLAN matching precedence	vlan precedence { mac-vlan ip-subnet-vlan }	Optional By default, VLANs are preferentially matched based on MAC addresses.

Protocol-Based VLAN Configuration

Introduction to Protocol-Based VLAN



Protocol-based VLANs are only applicable to Hybrid ports.

In this approach, inbound packets are assigned with different VLAN IDs based on their protocol type and encapsulation format. The protocols that can be used to categorize VLANs include: IP, IPX, and AppleTalk (AT). The encapsulation formats include: Ethernet II, 802.3 raw, 802.2 LLC, and 802.2 SNAP.

A protocol-based VLAN can be defined by a protocol template, which is determined by encapsulation format and protocol type. A port can be associated to multiple protocol templates. An untagged packet (that is, packet carrying no VLAN tag) reaching a port associated with a protocol-based VLAN will be processed as follows.

- If the packet matches a protocol template, the packet will be tagged with the VLAN ID of the protocol-based VLAN defined by the protocol template.
- If the packet matches no protocol template, the packet will be tagged with the default VLAN ID of the port.

The port processes a tagged packet (that is, a packet carrying a VLAN tag) in the same way as it processes packets of a port-based VLAN.

- If the port is configured to permit the VLAN identified by this VLAN tag, the port forwards the packet.
- If the port is configured to deny the VLAN identified by this VLAN tag, the port discards the packet.

This feature is mainly used to bind the service type with VLAN for ease of management and maintenance.

Configuring a Protocol-Based VLAN

Follow these steps to configure a protocol-based VLAN:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter VLAN view	vlan <i>vlan-id</i>	Required If the specified VLAN does not exist, this command creates the VLAN and then enters its view.
Configure the protocol-based VLAN and specify the protocol template	protocol-vlan [<i>protocol-index</i>] { at ipv4 ipv6 ipx { ethernetii llc raw snap } mode { ethernetii etype <i>etype-id</i> llc { dsap <i>dsap-id</i> [ssap <i>ssap-id</i>] } snap etype <i>etype-id</i> }	Required
Exit the VLAN view	quit	Required
Enter Ethernet port view or port group view	interface <i>interface-type</i> <i>interface-number</i> port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Use either command In Ethernet port view, the subsequent configurations only apply to the current port; in port group view, the subsequent configurations apply to all ports in the port group
Configure the port link type as Hybrid	port link-type hybrid	Required
Allow the packets of protocol-based VLANs to pass through the current Hybrid port in untagged way (with the tags of the packets stripped)	port hybrid vlan <i>vlan-id-list</i> untagged	Required
Configure the association between the Hybrid port and the protocol-based VLAN	port hybrid protocol-vlan vlan <i>vlan-id</i> { <i>protocol-index</i> [to <i>protocol-end</i>] all }	Required



CAUTION:

- *At present, the AppleTalk-based protocol template cannot be associated with a port on an Switch 4800G.*
- *Do not configure both the dsap-id and ssap-id arguments in the **protocol-vlan** command as 0xe0 or 0xff when configuring the user-defined*

template for **llc** encapsulation. Otherwise, the encapsulation format of the matching packets will be the same as that of the **ipx llc** or **ipx raw** packets respectively.

- When you use the **mode** keyword to configure a user-defined protocol template, do not set **etype-id** in **ethernetii etype** **etype-id** to **0x0800**, **0x8137**, **0x809b**, or **0x86dd**. Otherwise, the encapsulation format of the matching packets will be the same as that of the IPv4, IPX, AppleTalk, and IPv6 packets respectively.
- Do not configure a VLAN as both a protocol-based VLAN and a voice VLAN. Because a protocol-based VLAN requires that the inbound packets on the Hybrid port are untagged packets, whereas the Hybrid port working in auto voice VLAN mode only supports to process tagged voice traffic. For more information, refer to “Voice VLAN Configuration” on page 99.

Configuring IP-Subnet-Based VLAN

Introduction

In this approach, VLANs are categorized based on the source IP addresses and the subnet masks of packets. After receiving an untagged packet from a port, the device identifies the VLAN the packet belongs to based on the source address contained in the packet, and then forwards the packet in the VLAN. This allows packets from a certain network segment or with certain IP addresses to be forwarded in a specified VLAN.

Configuring an IP-Subnet-Based VLAN



This feature is only applicable to Hybrid ports.

Follow these steps to configure an IP-subnet-based VLAN:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter VLAN view	vlan <i>vlan-id</i>	-
Configure the association between an IP subnet with the current VLAN	ip-subnet-vlan [<i>ip-subnet-index</i>] ip <i>ip-address</i> [<i>mask</i>]	Required The configured IP network segment or IP address cannot be a multicast network segment or a multicast address
Return to system view	quit	-
Enter Ethernet port view or port group view	Enter Ethernet port view interface <i>interface-type</i> <i>interface-number</i> Enter port group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Use either command; In Ethernet port view, the subsequent configurations only apply to the current port; in port group view, the subsequent configurations apply to all ports in the port group
Configure port link type as Hybrid	port link-type hybrid	Required

To do...	Use the command...	Remarks
Allow an IP-subnet-based VLAN to pass through the current Hybrid port	port hybrid vlan <i>vlan-id-list</i> { tagged untagged }	Required
Configure the association between the Hybrid port and the IP-subnet-based VLAN	port hybrid ip-subnet-vlan vlan <i>vlan-id</i>	Required

Displaying and Maintaining VLAN

To do...	Use the command...	Remarks
Display the information about specific VLANs	display vlan [<i>vlan-id1</i> [to <i>vlan-id2</i>]] all dynamic reserved static]	Available in any view
Display the information about a VLAN interface	display interface Vlan-interface [<i>vlan-interface-id</i>]	Available in any view
Display all the ports with MAC address-based VLAN enabled.	display mac-vlan interface	Available in any view
Display the information about specific MAC address-to-VLAN entries	display mac-vlan { all dynamic mac-address <i>mac-addr</i> [mask <i>mac-mask</i>] } static vlan <i>vlan-id</i> }	Available in any view
Display the protocol information and protocol indexes of specified VLANs	display protocol-vlan vlan { <i>vlan-id</i> [to <i>vlan-id</i>] } all }	Available in any view
Display protocol-based VLAN information on specified interfaces	display protocol-vlan interface { <i>interface-type</i> <i>interface-number</i> [to <i>interface-type</i> <i>interface-number</i>] } all }	Available in any view
Display the IP-subnet-based VLAN information and IP subnet indexes of specified VLANs	display ip-subnet-vlan vlan { <i>vlan-id</i> [to <i>vlan-id</i>] } all }	Available in any view
Display the IP-subnet-based VLAN information and IP subnet index of specified ports	display ip-subnet-vlan interface { <i>interface-type</i> <i>interface-number</i> [to <i>interface-type</i> <i>interface-number</i>] } all }	Available in any view
Clear the statistics on a VLAN interface	reset counters interface Vlan-interface [<i>vlan-interface-id</i>]	Available in user view

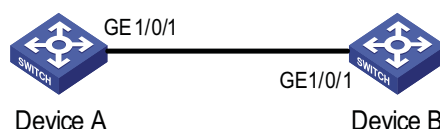
VLAN Configuration Example

Network requirements

- Device A connects to Device B through Trunk port GigabitEthernet 1/0/1;
- The default VLAN ID of the port is 100;
- This port allows packets from VLAN 2, VLAN 6 through VLAN 50, and VLAN 100 to pass through.

Network diagram

Figure 28 Network diagram for port-based VLAN configuration



Configuration procedure**1** Configure Device A

Create VLAN 2, VLAN 6 through VLAN 50, and VLAN 100.

```
<DeviceA> system-view
[DeviceA] vlan 2
[DeviceA-vlan2] quit
[DeviceA] vlan 100
[DeviceA-vlan100] vlan 6 to 50
Please wait... Done.
```

Enter GigabitEthernet 1/0/1 port view.

```
[DeviceA] interface GigabitEthernet 1/0/1
```

Configure GigabitEthernet 1/0/1 as a Trunk port and configure its default VLAN ID as 100.

```
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
[DeviceA-GigabitEthernet1/0/1] port trunk pvid vlan 100
```

Configure GigabitEthernet 1/0/1 to deny the packets of VLAN 1 (by default, the packets of VLAN 1 are permitted on all the ports).

```
[DeviceA-GigabitEthernet1/0/1] undo port trunk permit vlan 1
```

Configure packets from VLAN 2, VLAN 6 through VLAN 50, and VLAN 100 to pass through GigabitEthernet 1/0/1.

```
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan 2 6 to 50 100
Please wait... Done.
```

1 Configure Device B following similar steps as that of Device A.**Verification**

Verifying the configuration of Device A is similar to that of Device B. So only Device A is taken for example here.

Display the information about GigabitEthernet 1/0/1 of Device A to verify the above configurations.

```
<DeviceA> display interface GigabitEthernet 1/0/1
GigabitEthernet1/0/1 current state: UP
 IP Packet Frame Type: PKTFMT_ETHNT_2, Hardware Address: 0011-2233-5577
Description: GigabitEthernet1/0/1 Interface
Loopback is not set
Media type is twisted pair
Port hardware type is 1000_BASE_T
1000Mbps-speed mode, full-duplex mode
Link speed type is autonegotiation, link duplex type is autonegotiation
Flow-control is not enabled
The Maximum Frame Length is 9212
Broadcast MAX-ratio: 100%
Unicast MAX-ratio: 100%
Multicast MAX-ratio: 100%
Allow jumbo frame to pass
PVID: 100
Mdi type: auto
```

```

Link delay is 0(sec)
Port link-type: trunk
  Tagged   VLAN ID : 2, 6-50, 100
  Untagged VLAN ID : 2, 6-50, 100
Port priority: 0
Last 300 seconds input:  8 packets/sec 1513 bytes/sec  0%
Last 300 seconds output: 1 packets/sec 179 bytes/sec  0%
Input (total): 25504971 packets, 13911485028 bytes
                14288575 broadcasts, 11111535 multicasts
Input (normal): 25504971 packets, - bytes
                14288575 broadcasts, 11111535 multicasts
Input: 0 input errors, 0 runts, 0 giants, 0 throttles
       0 CRC, 0 frame, - overruns, 0 aborts
       - ignored, - parity errors
Output (total): 175995 packets, 31290143 bytes
                47 broadcasts, 68494 multicasts, 0 pauses
Output (normal): 175995 packets, - bytes
                47 broadcasts, 68494 multicasts, 0 pauses
Output: 0 output errors, - underruns, - buffer failures
        0 aborts, 0 deferred, 0 collisions, 0 late collisions
        0 lost carrier, - no carrier

```

The output above shows that:

- The port is a Trunk port (Port link-type: trunk).
- The default VLAN is VLAN 100 (PVID: 100).
- The port permits packets of VLAN 2, VLAN 6 through VLAN 50, and VLAN 100 (VLAN permitted: 2, 6-50, 100).

So the configuration is successful.

10

VOICE VLAN CONFIGURATION

When configuring Voice VLAN, go to these sections for information you are interested in:

- "Introduction to Voice VLAN" on page 99
- "Configuring Voice VLAN" on page 101
- "Displaying and Maintaining Voice VLAN" on page 103
- "Voice VLAN Configuration Examples" on page 103

Introduction to Voice VLAN

A voice VLAN is configured specially for voice traffic. By adding the ports that connect voice devices to the voice VLAN, you can configure quality of service (QoS for short) attributes for the voice traffic, improving transmission priority and ensuring voice quality. A device determines whether a received packet is a voice packet by checking its source MAC address. Packets containing source MAC addresses that comply with the voice device Organizationally Unique Identifier (OUI for short) addresses are regarded as voice traffic, and are forwarded to the voice VLAN.

You can configure the OUI addresses in advance or use the default OUI addresses, which are listed as follows.

Table 23 The default OUI addresses of different vendors

Number	OUI address	Vendors
1	0001-e300-0000	Siemens phone
2	0003-6b00-0000	Cisco phone
3	0004-0d00-0000	Avaya phone
4	0060-b900-0000	Philips/NEC phone
5	00d0-1e00-0000	Pingtel phone
6	00e0-7500-0000	Polycom phone
7	00e0-bb00-0000	3Com phone



- *As the first 24 bits of a MAC address (in binary format), an OUI address is a globally unique identifier assigned to a vendor by IEEE (Institute of Electrical and Electronics Engineers).*
- *You can add or remove default OUI address manually.*

Voice VLAN Modes on a Port

There are two voice VLAN modes on a port: automatic and manual (the mode here refers to the way of adding a port to a voice VLAN).

- In automatic mode, the system identifies the source MAC address contained in the protocol packets (untagged packets) sent when the IP phone is powered on

and matches it against the OUI addresses. If a match is found, the system will automatically add the port into the Voice VLAN and apply ACL rules and configure the packet precedence. An aging time can be configured for the voice VLAN. The system will remove a port from the voice VLAN if no voice packet is received from it after the aging time. The adding and removing of ports are automatically realized by the system.

- In manual mode, administrators add the IP phone access port to the voice VLAN manually. It then identifies the source MAC address contained in the packet, matches it against the OUI addresses. If a match is found, the system issues ACL rules and configures the precedence for the packets. In this mode, the operation of adding ports to and removing ports from the voice VLAN are carried out by the administrators.
- Both modes forward tagged packets according to their tags.

The following table lists the co-relation between the port voice VLAN mode, the voice traffic type of an IP phone, and the port link type.

Table 24 Voice VLAN operating mode and the corresponding voice traffic types

Port voice VLAN mode	Voice traffic type	Port link type
Automatic mode	Tagged voice traffic	Access: not supported
		Trunk: supported provided that the default VLAN of the access port exists and is not the voice VLAN and that the access port belongs to the voice VLAN Hybrid: supported provided that the default VLAN of the access port exists and is not the voice VLAN, and is in the list of tagged VLANs whose packets can pass through the access port
Manual mode	Untagged voice traffic	Access, Trunk, Hybrid: not supported
	Tagged voice traffic	Access: not supported
		Trunk: supported provided that the default VLAN of the access port exists and is not the voice VLAN and that the access port belongs to the default VLAN Hybrid: supported provided that the default VLAN of the access port exists and is not the voice VLAN, and is in the list of tagged VLANs whose packets can pass through the access port
		Untagged voice traffic

**CAUTION:**

- If the voice traffic sent by an IP phone is tagged and that the access port has 802.1x authentication and Guest VLAN enabled, assign different VLAN IDs for the voice VLAN, the default VLAN of the access port, and the 802.1x guest VLAN.
- If the voice traffic sent by an IP phone is untagged, to realize the voice VLAN feature, the default VLAN of the access port can only be configured as the voice VLAN. Note that at this time 802.1x authentication function cannot be realized.



- The default VLAN for all ports is VLAN 1. Using commands, users can either configure the default VLAN of a port, or configure to allow a certain VLAN to pass through the port. For more information, refer to section "Port-Based VLAN Configuration" on page 87.
- Use the **display interface** command to display the default VLAN and the VLANs that are allowed to go through a certain port.

Security Mode and Normal Mode for the Voice VLAN

Voice VLAN modes fall into security mode and normal mode based on the filtering mechanisms of the voice VLAN-enabled ports on the inbound packets. In the two modes, the voice VLAN-enabled ports process untagged packets and packets with the voice VLAN tags in different ways, as shown in the following table:

Voice VLAN mode	Inbound packet type	Processing way
Security mode	Untagged packets	If the source MAC addresses of the packets are OUI addresses that can be identified by the system, send the packets to the voice VLAN; otherwise, discard the packets.
	Packets with the voice VLAN tag	
Normal mode	Untagged packets	The packet source MAC address will not be checked, and all packets can be transmitted in the voice VLAN.
	Packets with the voice VLAN tag	

In the two modes, the port processes a packet with other VLAN tag in the same way, that is, forwards the packet if the VLAN is allowed on the port, or discards the packet if the VLAN is not allowed on the port.

It is recommended that you do not mix voice packets with other types of data in a voice VLAN. If necessary, please ensure that the security mode is disabled.

Configuring Voice VLAN

Configuration Prerequisites

- Create the corresponding VLAN before configuring the voice VLAN;
- As a default VLAN, VLAN 1 does not need to be created. However, it cannot be enabled with the voice VLAN feature.

Configuring Voice VLAN Mode on a Port to Automatic Mode

Follow these steps to set the port voice VLAN mode to automatic:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Configure the aging time of the voice VLAN	voice vlan aging <i>minutes</i>	Optional Only applicable to ports in automatic mode and defaults to 1,440 minutes
Enable the security mode for the voice VLAN	voice vlan security enable	Optional Enabled by default
Configure the OUI address for the voice VLAN	voice vlan mac-address <i>oui mask oui-mask [description text]</i>	Optional By default, each voice VLAN has default OUI addresses configured. Refer to Table 23 for the default OUI addresses of different vendors.
Enable the voice VLAN feature globally	voice vlan <i>vlan-id</i> enable	Required
Enter Ethernet port view	interface <i>interface-type interface-number</i>	-
Configure the port voice VLAN mode as automatic	voice vlan mode auto	Optional Automatic mode by default Different voice VLAN modes can be configured on different ports, independent of one another.
Enable the voice VLAN feature on the port	voice vlan enable	Required Not enabled by default



- *Do not configure a VLAN as both a protocol-based VLAN and a voice VLAN. Because a protocol-based VLAN requires that the inbound packets on the Hybrid port are untagged packets (refer to section “Protocol-Based VLAN Configuration” on page 92), whereas the Hybrid port working in auto voice VLAN mode only supports to process tagged voice traffic.*
- *The default VLAN of a port in automatic mode cannot be configured as the voice VLAN. Otherwise, the system will prompt error information.*

Configuring Voice VLAN Mode on a Port to Manual Mode

Follow these steps to set the port voice VLAN mode to manual:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the security mode of a voice VLAN	voice vlan security enable	Optional Enabled by default
Configure the OUI address of a voice VLAN	voice vlan mac-address <i>oui mask oui-mask [description text]</i>	Optional By default, each voice VLAN has default OUI addresses configured. Refer to Table 23 for the default OUI addresses of different vendors.
Enable the voice VLAN feature globally	voice vlan <i>vlan-id</i> enable	Required
Enter Ethernet port view	interface <i>interface-type interface-number</i>	-

To do...		Use the command...	Remarks
Configure the working mode as manual		undo voice vlan mode auto	Required Disabled by default
Add the ports in manual mode to the voice VLAN	Access port	Refer to "Configuring an Access-Port-Based VLAN" on page 88.	Use one of the three approaches.
	Trunk port	Refer to "Configuring a Trunk-Port-Based VLAN" on page 89.	After you add an Access port to the voice VLAN, the voice VLAN becomes the default VLAN of the port automatically.
	Hybrid port	Refer to "Configuring a Hybrid-Port-Based VLAN" on page 90.	
Configure the voice VLAN as the default VLAN of the port	Trunk port	Refer to section "Configuring a Trunk-Port-Based VLAN" on page 89	Optional This operation is required if the inbound voice traffic is untagged. If the inbound voice traffic is tagged, do not configure the voice VLAN as the default VLAN of the port.
	Hybrid port	Refer to "Configuring a Hybrid-Port-Based VLAN" on page 90.	
Enable the voice VLAN feature on the port		voice vlan enable	Required



- Only one VLAN of a device can have the voice VLAN function enabled at a time, and the VLAN must be an existing static VLAN.
- A port that is in a link aggregation port group cannot have the voice VLAN feature enabled.
- If a port is enabled with voice VLAN and works in the manual voice VLAN mode, you need to add the port to the voice VLAN manually to make the voice VLAN takes effect on the port.

Displaying and Maintaining Voice VLAN

To do...	Use the command...	Remarks
Display the voice VLAN state	display voice vlan state	Available in any view
Display the OUI addresses currently supported by system	display voice vlan oui	Available in any view

Voice VLAN Configuration Examples

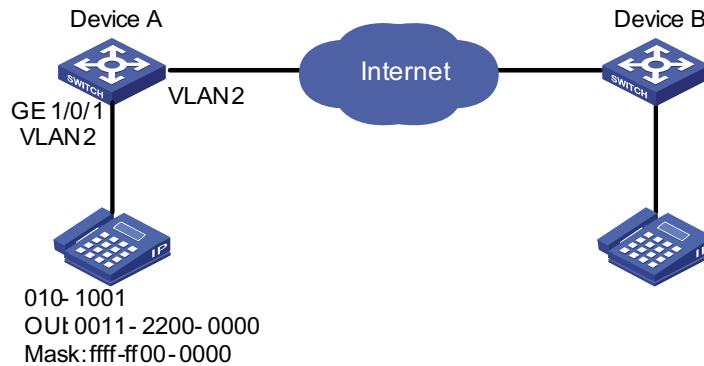
Automatic Voice VLAN Mode Configuration Example

Network requirement

- Create VLAN 2 and configure it as a voice VLAN with an aging time of 100 minutes.
- The voice traffic sent by the IP phones is tagged. Configure GigabitEthernet 1/0/1 as a Hybrid port and as the access port, with VLAN 6 as the default VLAN.
- The device allows voice packets from GigabitEthernet 1/0/1 with an OUI address of 0011-2200-0000 and a mask of ffff-ff00-0000 to be forwarded through the voice VLAN.

Network diagram

Figure 29 Network diagram for automatic voice VLAN mode configuration



Configuration procedure

Create VLAN 2 and VLAN 6.

```
<DeviceA> system-view
[DeviceA] vlan 2
[DeviceA-vlan2] quit
[DeviceA] vlan 6
[DeviceA-vlan6] quit
```

Configure the voice VLAN aging time.

```
[DeviceA] voice vlan aging 100
```

Configure the OUI address 0011-2200-0000 as the legal address of the voice VLAN.

```
[DeviceA] voice vlan mac-address 0011-2200-0000 mask ffff-ff00-0000
```

Enable the voice VLAN feature globally.

```
[DeviceA] voice vlan 2 enable
```

Configure the voice VLAN mode on GigabitEthernet 1/0/1 as automatic. (Optional, by default, the voice VLAN mode on a port is automatic mode)

```
[DeviceA] interface GigabitEthernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] voice vlan mode auto
```

Configure GigabitEthernet 1/0/1 as a Hybrid port.

```
[DeviceA-GigabitEthernet1/0/1] port link-type access
Please wait... Done.
[DeviceA-GigabitEthernet1/0/1] port link-type hybrid
```

Configure the default VLAN of the port as VLAN 6 and allow packets from VLAN 6 to pass through the port.

```
[DeviceA-GigabitEthernet1/0/1] port hybrid pvid vlan 6
[DeviceA-GigabitEthernet1/0/1] port hybrid vlan 6 tagged
```

Enable the voice VLAN feature on the port.

```
[DeviceA-GigabitEthernet1/0/1] voice vlan enable
[DeviceA-GigabitEthernet1/0/1] return
```

Verification

Display information about the OUI addresses, OUI address masks, and descriptive strings.

```
<DeviceA> display voice vlan oui
Oui Address      Mask              Description
0001-e300-0000   ffff-ff00-0000   Siemens phone
0003-6b00-0000   ffff-ff00-0000   Cisco phone
0004-0d00-0000   ffff-ff00-0000   Avaya phone
0011-2200-0000   ffff-ff00-0000
0060-b900-0000   ffff-ff00-0000   Philips/NEC phone
00d0-1e00-0000   ffff-ff00-0000   Pingtel phone
00e0-7500-0000   ffff-ff00-0000   Polycom phone
00e0-bb00-0000   ffff-ff00-0000   3com phone
```

Display the current Voice VLAN state.

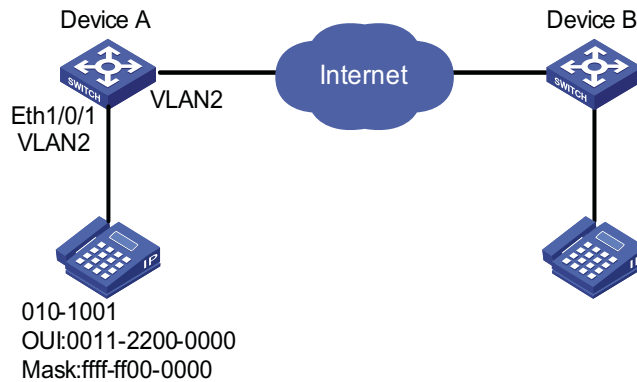
```
<DeviceA> display voice vlan state
Voice VLAN status: ENABLE
Voice VLAN ID: 2
Voice VLAN security mode: Security
Voice VLAN aging time: 100 minutes
Voice VLAN enabled port and its mode:
PORT              MODE
-----
GigabitEthernet1/0/1      AUTO

<DeviceA>
```

Manual Voice VLAN Mode Configuration Example

Network requirement

- Create VLAN 2 and configure it as a voice VLAN.
- The voice traffic sent by the IP phones is untagged. Configure GigabitEthernet 1/0/1 as a Hybrid port and as the access port.
- GigabitEthernet 1/0/1 works in manual mode. It only allows voice packets with an OUI address of 0011-2200-0000, a mask of ffff-ff00-0000, and a descriptive string of **test** to be forwarded through the voice VLAN.

Network diagram**Figure 30** Network diagram for manual voice VLAN mode configuration**Configuration procedure**

Configure the voice VLAN to work in security mode and only allows legal voice packets to pass through the voice VLAN enabled port. (Optional, enabled by default)

```
<DeviceA> system-view
[DeviceA] voice vlan security enable
```

Configure the OUI address 0011-2200-0000 as the legal voice VLAN address.

```
[DeviceA] voice vlan mac-address 0011-2200-0000 mask ffff-ff00-0000 description test
```

Create VLAN 2. Enable voice VLAN feature for it.

```
[DeviceA] vlan 2
[DeviceA-vlan2] quit
[DeviceA] voice vlan 2 enable
```

Configure GigabitEthernet 1/0/1 to work in manual mode.

```
[DeviceA] interface GigabitEthernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] undo voice vlan mode auto
```

Configure GigabitEthernet 1/0/1 as a Hybrid port.

```
[DeviceA-GigabitEthernet1/0/1]port link-type access
Please wait... Done.
[DeviceA-GigabitEthernet1/0/1]port link-type hybrid
```

Configure the default VLAN of GigabitEthernet 1/0/1 as voice VLAN and add the voice VLAN to the list of tagged VLANs whose packets can pass through the port.

```
[DeviceA-GigabitEthernet1/0/1] port hybrid pvid vlan 2
[DeviceA-GigabitEthernet1/0/1] port hybrid vlan 2 untagged
```

Enable the voice VLAN feature of GigabitEthernet 1/0/1.

```
[DeviceA-GigabitEthernet1/0/1] voice vlan enable
```

Verification

Display information about the OUI addresses, OUI address masks, and descriptive strings.

```
<DeviceA> display voice vlan oui
Oui Address      Mask             Description
0001-e300-0000   ffff-ff00-0000  Siemens phone
0003-6b00-0000   ffff-ff00-0000  Cisco phone
0004-0d00-0000   ffff-ff00-0000  Avaya phone
0011-2200-0000   ffff-ff00-0000  test
0060-b900-0000   ffff-ff00-0000  Philips/NEC phone
00d0-1e00-0000   ffff-ff00-0000  Pingtel phone
00e0-7500-0000   ffff-ff00-0000  Polycom phone
00e0-bb00-0000   ffff-ff00-0000  3com phone
```

Display the current voice VLAN state.

```
<DeviceA> display voice vlan state
Voice VLAN status: ENABLE
Voice VLAN ID: 2
Voice VLAN security mode: Security
Voice VLAN aging time: 100 minutes
Voice VLAN enabled port and its mode:
PORT                MODE
-----
GigabitEthernet1/0/1      MANUAL
```


11

GVRP CONFIGURATION

GARP VLAN Registration Protocol (GVRP) is a GARP application. It functions based on the operating mechanism of GARP to maintain and propagate dynamic VLAN registration information for the GVRP devices on the network.

When configuring GVRP, go to these sections for information you are interested in:

- "Introduction to GVRP" on page 109
- "GVRP Configuration Task List" on page 112
- "Configuring GVRP" on page 112
- "Displaying and Maintaining GVRP" on page 114
- "GVRP Configuration Examples" on page 114

Introduction to GVRP

GARP Generic Attribute Registration Protocol (GARP) provides a mechanism that allows participants in a GARP application to distribute, propagate, and register with other participants in a bridged LAN the attributes specific to the GARP application, such as the VLAN or multicast address attribute.

GARP itself does not exist on a device as an entity. GARP-compliant participants are known as GARP applications. One example is GVRP. When a GARP participant is present on a port on your device, the port is regarded as a GARP participant.

GARP messages and timers

1 GARP messages

GARP participants exchange information through the following three types of messages: Join message, Leave message, and LeaveAll message.

- A GARP participant uses Join messages to have its attributes registered on other devices. A GARP participant also sends Join messages to register attributes on other GARP participants when it receives Join messages from other GARP participants or static attributes are configured on it.
- A GARP participant uses Leave messages to have its attributes deregistered on other devices. A GARP participant also sends Leave messages when it receives Leave messages from other GARP participants or static attributes are deregistered on it.
- LeaveAll messages are used to deregister all the attributes, through which all the other GARP participants begin to have all their attributes registered. A

GARP participant sends LeaveAll messages upon the expiration of the LeaveAll timer, which is triggered when the GARP participant is created.

Join messages, Leave messages, and LeaveAll message make sure the re-registration and deregistration of GARP attributes are performed in an orderly way.

Through message exchange, all attribute information that needs registration propagates to all GARP participants throughout a LAN.

2 GARP timers

The interval of sending of GARP messages is controlled by the following four timers:

- Hold timer -- A GARP participant usually does not forwards a received registration request immediately after it receives a registration request, instead, it waits for the expiration of the hold timer. That is, a GARP participant sends Join messages when the hold timer expires. The Join message contains all the registration information received during the latest Hold timer cycle. Such a mechanism saves the bandwidth.
- Join timer -- Each GARP participant sends a Join message twice for reliability sake and uses a join timer to set the sending interval. If the first Join message is not acknowledged after the interval defined by the Join timer, the GARP participant sends the second Join message.
- Leave timer -- Starts upon receipt of a Leave message sent for deregistering some attribute information. If no Join message is received before this timer expires, the GARP participant removes the attribute information as requested.
- LeaveAll timer -- Starts when a GARP participant starts. When this timer expires, the entity sends a LeaveAll message so that other participants can re-register its attribute information. Then, a LeaveAll timer starts again.



- *The settings of GARP timers apply to all GARP applications, such as GVRP, on a LAN.*
- *Unlike other three timers, which are set on a port basis, the LeaveAll timer is set in system view and takes effect globally.*
- *A GARP participant may send LeaveAll messages at the interval set by its LeaveAll timer or the LeaveAll timer on another device on the network, whichever is smaller. This is because each time a device on the network receives a LeaveAll message it resets its LeaveAll timer.*

Operating mechanism of GARP

The GARP mechanism allows the configuration of a GARP participant to propagate throughout a LAN quickly. In GARP, a GARP participant registers or deregisters its attributes with other participants by making or withdrawing declarations of attributes and at the same time, based on received declarations or withdrawals, handles attributes of other participants. When a port receives an attribute declaration, it registers the attribute; when a port receives an attribute withdrawal, it deregisters the attribute.

GARP participants send protocol data units (PDU) with a particular multicast MAC address as destination. Based on this address, a device can identify to which GVRP application, GVRP for example, should a GARP PDU be delivered.

GARP message format

The following figure illustrates the GARP message format.

Figure 31 GARP message format

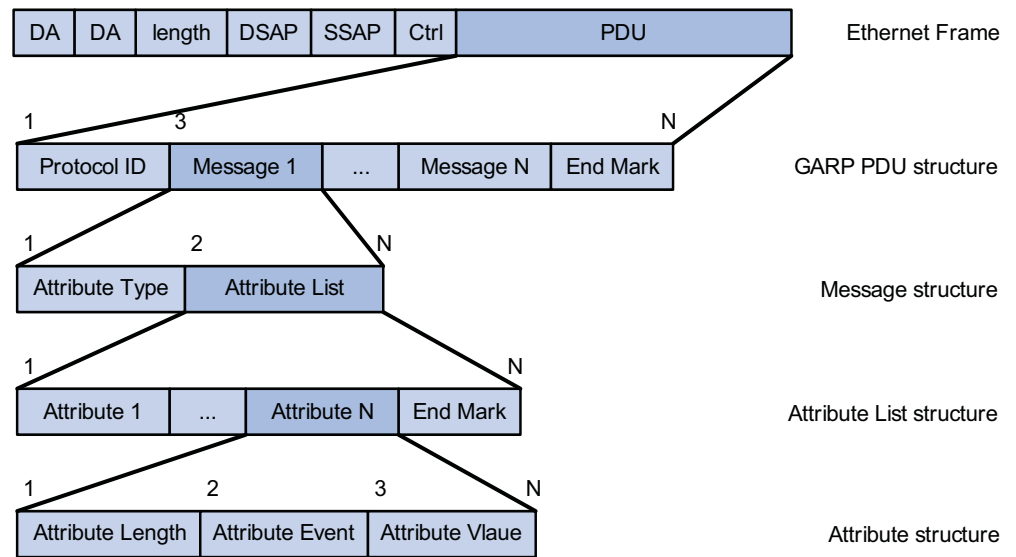


Table 25 describes the GARP message fields.

Table 25 Description on the GARP message fields

Field	Description	Value
Protocol ID	Protocol identifier for GARP	1
Message	One or multiple messages, each containing an attribute type and an attribute list	--
Attribute Type	Defined by the concerned GARP application	0x01 for GVRP, indicating the VLAN ID attribute
Attribute List	Contains one or multiple attributes	--
Attribute	Consists of an Attribute Length, an Attribute Event, and an Attribute Value	--
Attribute Length	Number of octets occupied by an attribute, inclusive of the attribute length field	2 to 255 (in bytes)
Attribute Event	Event described by the attribute	0: LeaveAll event 1: JoinEmpty event 2: JoinIn event 3: LeaveEmpty event 4: LeaveIn event 5: Empty event

Table 25 Description on the GARP message fields

Field	Description	Value
Attribute Value	Attribute value	VLAN ID for GVRP If the Attribute Event is LeaveAll, Attribute Value is omitted.
End Mark	Indicates the end of a GARP PDU	0x00

GVRP GVRP enables a device to propagate local VLAN registration information to other participant devices and dynamically update the VLAN registration information from other devices to its local database about active VLAN members and through which port they can be reached. It thus ensures that all GVRP participants on a bridged LAN maintain the same VLAN registration information. The VLAN registration information propagated by GVRP includes both manually configured local static entries and dynamic entries from other devices.

GVRP provides the following three registration types on a port:

- Normal -- Enables the port to dynamically register and deregister VLANs, and to propagate both dynamic and static VLAN information.
- Fixed -- Disables the port to dynamically register and deregister VLANs or propagate information about dynamic VLANs, but allows the port to propagate information about static VLANs. A trunk port with fixed registration type thus allows only manually configured VLANs to pass through even though it is configured to carry all VLANs.
- Forbidden -- Disables the port to dynamically register and deregister VLANs, and to propagate VLAN information except information about VLAN 1. A trunk port with forbidden registration type thus allows only VLAN 1 to pass through even though it is configured to carry all VLANs.

Protocols and Standards GVRP is described in IEEE 802.1Q.

GVRP Configuration Task List



GVRP can only be configured on Trunk ports.

Complete the following tasks to configure GVRP:

Task	Remarks
"Enabling GVRP" on page 112	Required
"Configuring GARP Timers" on page 113	Optional

Configuring GVRP

Enabling GVRP Follow these steps to enable GVRP on a trunk port:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enable GVRP globally	gvrp	Required Globally disabled by default
Enter Ethernet port view or port-group view	Enter Ethernet port view interface <i>interface-type</i> <i>interface-number</i> Enter port group view port-group { aggregation <i>agg-id</i> manual <i>port-group-name</i> }	Use either command. In Ethernet port view, the subsequent configurations only apply to the current port; in port group view, the subsequent configurations apply to all ports in the port group.
Enable GVRP on the port	gvrp	Required Disabled by default
Configure the GVRP registration mode on the port	gvrp registration { fixed forbidden normal }	Optional The default is normal .



Because GVRP is not compatible with the BPDU tunneling feature, you must disable BPDU tunneling before enabling GVRP on a BPDU tunneling-enabled Ethernet port.

Configuring GARP Timers

Follow these steps to configure GARP timers:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Configure the GARP LeaveAll timer	garp timer leaveall <i>timer-value</i>	Optional The default is 1000 centiseconds.
Enter Ethernet port view or port-group view	Enter Ethernet port view interface <i>interface-type</i> <i>interface-number</i> Enter port-group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Use either command. In Ethernet port view, the subsequent configurations only apply to the current port; in port group view, the subsequent configurations apply to all ports in the port group.
Configure the hold timer, join timer, and leave timer	garp timer { hold join leave } <i>timer-value</i>	Optional The default is 10 centiseconds for the hold timer, 20 centiseconds for the join timer, and 60 centiseconds for the leave timer.

As for the GARP timers, note that:

- The setting of each timer must be a multiple of five (in centiseconds).
- The settings of the timers are correlated. If you fail to set a timer to a certain value, you can try to adjust the settings of the rest timers. Table 26 shows the relationship of the timers.

Table 26 Dependencies of GARP timers

Timer	Lower limit	Upper limit
Hold	10 centiseconds	Not greater than half of the join timer setting
Join	Not less than two times the hold timer setting	Less than half of the leave timer setting
Leave	Greater than two times the join timer setting	Less than the LeaveAll timer setting
LeaveAll	Greater than the leave timer setting	32765 centiseconds

Displaying and Maintaining GVRP

To do...	Use the command...	Remarks
Display statistics about GARP	display garp statistics [interface <i>interface-list</i>]	Available in any view
Display GARP timers for specified or all ports	display garp timer [interface <i>interface-list</i>]	Available in any view
Display the local VLAN information maintained by GVRP	display gvrp local-vlan interface <i>interface-type</i> <i>interface-number</i>	Available in any view
Display the current GVRP state	display gvrp state interface <i>interface-type interface-number</i> vlan <i>vlan-id</i>	Available in any view
Display statistics about GVRP	display gvrp statistics [interface <i>interface-list</i>]	Available in any view
Display the global GVRP state	display gvrp status	Available in any view
Display the information about dynamic VLAN operations performed on a port	display gvrp vlan-operation interface <i>interface-type</i> <i>interface-number</i>	Available in any view
Clear the GARP statistics	reset garp statistics [interface <i>interface-list</i>]	Available in user view

GVRP Configuration Examples

GVRP Configuration Example 1

Network requirements

Configure GVRP for dynamic VLAN information registration and update among devices, adopting the normal registration mode on ports.

Network diagram

Figure 32 Network diagram for GVRP configuration



Configuration procedure**1** Configure Device A

Enable GVRP globally.

```
<DeviceA> system-view
[DeviceA] gvrp
```

Configure port GigabitEthernet 1/0/1 as a Trunk port, allowing all VLANs to pass.

```
[DeviceA] interface GigabitEthernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan all
```

Enable GVRP on GigabitEthernet 1/0/1, the Trunk port.

```
[DeviceA-GigabitEthernet1/0/1] gvrp
[DeviceA-GigabitEthernet1/0/1] quit
```

Create VLAN 2 (a static VLAN).

```
[DeviceA] vlan 2
```

2 Configure Device B

Enable GVRP globally.

```
<DeviceB> system-view
[DeviceB] gvrp
```

Configure port GigabitEthernet 1/0/1 as a Trunk port, allowing all VLANs to pass.

```
[DeviceB] interface GigabitEthernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan all
```

Enable GVRP on GigabitEthernet 1/0/1, the Trunk port.

```
[DeviceB-GigabitEthernet1/0/1] gvrp
[DeviceB-GigabitEthernet1/0/1] quit
```

Create VLAN 3 (a static VLAN).

```
[DeviceB] vlan 3
```

3 Verify the configuration

Display dynamic VLAN information on Device A.

```
[DeviceA] display vlan dynamic
Now, the following dynamic VLAN exist(s):
 3
```

Display dynamic VLAN information on Device B.

```
[DeviceB] display vlan dynamic
Now, the following dynamic VLAN exist(s):
 2
```

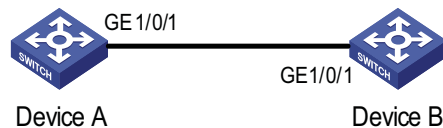
GVRP Configuration Example II

Network requirements

Configure GVRP for dynamic VLAN information registration and update among devices. Specify fixed GVRP registration on Device A and normal GVRP registration on Device B.

Network diagram

Figure 33 Network diagram for GVRP configuration



Configuration procedure

1 Configure Device A

Enable GVRP globally.

```
<DeviceA> system-view
[DeviceA] gvrp
```

Configure port GigabitEthernet 1/0/1 as a Trunk port, allowing all VLANs to pass.

```
[DeviceA] interface GigabitEthernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan all
```

Enable GVRP on GigabitEthernet 1/0/1.

```
[DeviceA-GigabitEthernet1/0/1] gvrp
```

Set the GVRP registration type to fixed on the port.

```
[DeviceA-GigabitEthernet1/0/1] gvrp registration fixed
[DeviceA-GigabitEthernet1/0/1] quit
```

Create VLAN 2 (a static VLAN).

```
[DeviceA] vlan 2
```

2 Configure Device B

Enable GVRP globally.

```
<DeviceB> system-view
[DeviceB] gvrp
```

Configure port GigabitEthernet 1/0/1 as a Trunk port, allowing all VLANs to pass.

```
[DeviceB] interface GigabitEthernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan all
```

Enable GVRP on GigabitEthernet 1/0/1.


```
[DeviceB-GigabitEthernet1/0/1] gvrp
[DeviceB-GigabitEthernet1/0/1] quit
```

Create VLAN 3 (a static VLAN).

```
[Sysname] vlan 3
```

3 Verify the configuration

Display dynamic VLAN information on Device A.

```
[DeviceA] display vlan dynamic
No dynamic vlans exist!
```

Display dynamic VLAN information on Device B.

```
[DeviceB] display vlan dynamic
Now, the following dynamic VLAN exist(s):
2
```

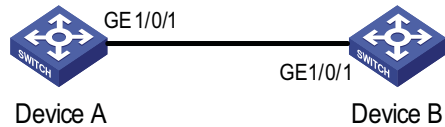
GVRP Configuration Example III

Network requirements

To prevent dynamic VLAN information registration and update among devices, set the GVRP registration mode to **forbidden** on Device A and **normal** on Device B.

Network diagram

Figure 34 Network diagram for GVRP configuration



Configuration procedure

1 Configure Device A

Enable GVRP globally.

```
<DeviceA> system-view
[DeviceA] gvrp
```

Configure port GigabitEthernet 1/0/1 as a Trunk port, allowing all VLANs to pass.

```
[DeviceA] interface GigabitEthernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] port link-type trunk
[DeviceA-GigabitEthernet1/0/1] port trunk permit vlan all
```

Enable GVRP on GigabitEthernet 1/0/1.

```
[DeviceA-GigabitEthernet1/0/1] gvrp
```

Set the GVRP registration type to forbidden on the port.

```
[DeviceA-GigabitEthernet1/0/1] gvrp registration forbidden
[DeviceA-GigabitEthernet1/0/1] quit
```

Create VLAN 2 (a static VLAN).

```
[DeviceA] vlan 2
```

2 Configure Device B

Enable GVRP globally.

```
<DeviceB> system-view
[DeviceB] gvrp
```

Configure port GigabitEthernet 1/0/1 as a Trunk port, allowing all VLANs to pass.

```
[DeviceB] interface GigabitEthernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] port link-type trunk
[DeviceB-GigabitEthernet1/0/1] port trunk permit vlan all
```

Enable GVRP on GigabitEthernet 1/0/1.

```
[DeviceB-GigabitEthernet1/0/1] gvrp
[DeviceB-GigabitEthernet1/0/1] quit
```

Create VLAN 3 (a static VLAN).

```
[DeviceB] vlan 3
```

3 Verify the configuration

Display dynamic VLAN information on Device A.

```
[DeviceA] display vlan dynamic
No dynamic vlans exist!
```

Display the VLANs allowed on GigabitEthernet 1/0/1.

```
[DeviceA] display interface GigabitEthernet 1/0/1
GigabitEthernet1/0/1 current state: DOWN
IP Packet Frame Type: PKTFMT_ETHNT_2, Hardware Address: 00e0-fc55-0010
Description: GigabitEthernet1/0/1 Interface
Loopback is not set
Media type is twisted pair
Port hardware type is 1000_BASE_T
Unknown-speed mode, unknown-duplex mode
Link speed type is autonegotiation, link duplex type is autonegotiation
Flow-control is not enabled
The Maximum Frame Length is 9212
Broadcast MAX-ratio: 100%
Unicast MAX-ratio: 100%
Multicast MAX-ratio: 100%
Allow jumbo frame to pass
PVID: 1
Mdi type: auto
Link delay is 0(sec)
Port link-type: trunk
  VLAN passing  : 1(default vlan)
  VLAN permitted: 1(default vlan)
(Omitted)
```

The above output indicates that port GigabitEthernet 1/0/1 only allows packets of VLAN 1 to pass.

Display dynamic VLAN information on Device B.

```
[DeviceB] display vlan dynamic  
No dynamic vlans exist!
```


12

IP ADDRESSING CONFIGURATION

When assigning IP addresses to interfaces on your device, go to these sections for information you are interested in:

- "IP Addressing Overview" on page 121
- "Configuring IP Addresses" on page 123
- "Displaying and Maintaining IP Addressing" on page 126

IP Addressing Overview

This section covers these topics:

- "IP Address Classes" on page 121
- "Special Case IP Addresses" on page 122
- "Subnetting and Masking" on page 122

IP Address Classes

IP addressing uses a 32-bit address to identify each host on a network. An example is 01010000100000001000000010000000 in binary. To make IP addresses in 32-bit form easier to read, they are written in dotted decimal notation, each being four octets in length, for example, 10.1.1.1 for the address just mentioned.

Each IP address breaks down into two parts:

- Net-id: First several bits of the IP address defining a network, also known as class bits.
- Host-id: Identifies a host on a network.

For administration sake, IP addresses are divided into five classes. Which class an IP address belongs to depends on the first one to four bits of the net-id, as shown in the following figure (in which the blue parts represent the address class).

Figure 35 IP address classes

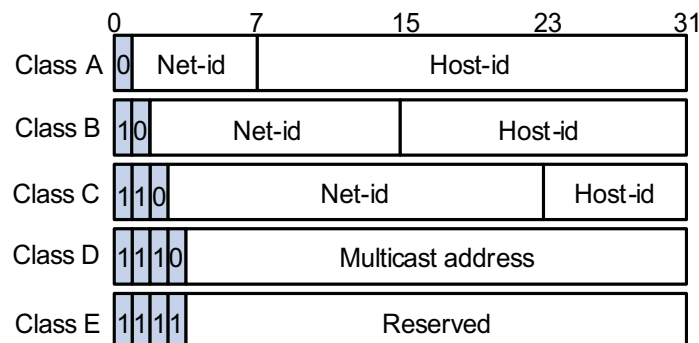


Table 27 describes the address ranges of these five classes. Currently, the first three classes of IP addresses are used in quantity.

Table 27 IP address classes and ranges

Class	Address range	Description
A	0.0.0.0 to 127.255.255.255	The IP address 0.0.0.0 is used by a host at bootstrap for temporary communication. This address is never a valid destination address. Addresses starting with 127 are reserved for loopback test. Packets destined to these addresses are processed locally as input packets rather than sent to the link.
B	128.0.0.0 to 191.255.255.255	--
C	192.0.0.0 to 223.255.255.255	--
D	224.0.0.0 to 239.255.255.255	Multicast address.
E	240.0.0.0 to 255.255.255.255	Reserved for future use except for the broadcast address 255.255.255.255.

Special Case IP Addresses

The following IP addresses are for special use, and they cannot be used as host IP addresses:

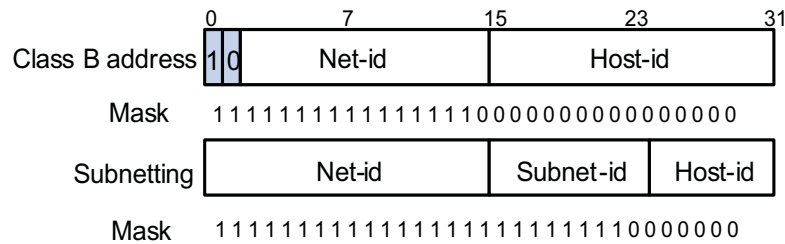
- IP address with an all-zero net ID: Identifies a host on the local network. For example, IP address 0.0.0.16 indicates the host with a host ID of 16 on the local network.
- IP address with an all-zero host ID: Identifies a network.
- IP address with an all-one host ID: Identifies a directed broadcast address. For example, a packet with the destination address of 192.168.1.255 will be broadcasted to all the hosts on the network 192.168.1.0.

Subnetting and Masking

Subnetting was developed to address the risk of IP address exhaustion resulting from fast expansion of the Internet. The idea is to break a network down into smaller networks called subnets by using some bits of the host-id to create a subnet-id. To identify the boundary between the host-id and the combination of net-id and subnet-id, masking is used. (When subnetting is not adopted, a mask identifies the boundary between the host-id and the host-id.)

Each subnet mask comprises 32 bits related to the corresponding bits in an IP address. In a subnet mask, the part containing consecutive ones identifies the combination of net-id and subnet-id whereas the part containing consecutive zeros identifies the host-id.

Figure 36 shows how a Class B network is subnetted.

Figure 36 Subnet a Class B network

While allowing you to create multiple logical networks within a single Class A, B, or C network, subnetting is transparent to the rest of the Internet. All these networks still appear as one. As subnetting adds an additional level, subnet-id, to the two-level hierarchy with IP addressing, IP routing now involves three steps: delivery to the site, delivery to the subnet, and delivery to the host.

In the absence of subnetting, some special addresses such as the addresses with the net-id of all zeros and the addresses with the host-id of all ones, are not assignable to hosts. The same is true of subnetting. When designing your network, you should note that subnetting is somewhat a tradeoff between subnets and accommodated hosts. For example, a Class B network can accommodate 65,534 ($2^{16} - 2$). Of the two deducted Class B addresses, one with an all-one host-id is the broadcast address and the other with an all-zero host-id is the network address) hosts before being subnetted. After you break it down into 512 (2^9) subnets by using the first 9 bits of the host-id for the subnet, you have only 7 bits for the host-id and thus have only 126 ($2^7 - 2$) hosts in each subnet. The maximum number of hosts is thus 64,512 (512×126), 1022 less after the network is subnetted.

Class A, B, and C networks, before being subnetted, use these default masks (also called natural masks): 255.0.0.0, 255.255.0.0, and 255.255.255.0 respectively.

Configuring IP Addresses

Besides directly assigning an IP address to an interface, you may configure the interface to obtain one through BOOTP or DHCP as alternatives. If you change the way an interface obtains an IP address, from manual assignment to BOOTP for example, the IP address obtained from BOOTP will overwrite the old one manually assigned.



This chapter only covers how to assign an IP address manually. For other approaches, refer to "DHCP Overview" on page 791.

This section includes:

- "Assigning an IP Address to an Interface" on page 123
- "IP Addressing Configuration Example" on page 124

Assigning an IP Address to an Interface

You may assign an interface multiple IP addresses, one primary and multiple secondaries, to connect multiple logical subnets on the same physical subnet.

Follow these steps to assign an IP address to an interface:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--
Assign an IP address to the interface	ip address <i>ip-address</i> { <i>mask</i> <i>mask-length</i> } [sub]	Required No IP address is assigned by default.

**CAUTION:**

- *The primary IP address you assigned to the interface can overwrite the old one if there is any.*
- *An interface cannot be configured with a secondary IP address if the interface has been configured to obtain an IP address through BOOTP or DHCP.*
- *The primary and secondary IP addresses you assign to the interface can be located on the same network segment. However, this should not violate the rule that different physical interfaces on your device must reside on different network segments.*

IP Addressing Configuration Example

Network requirements

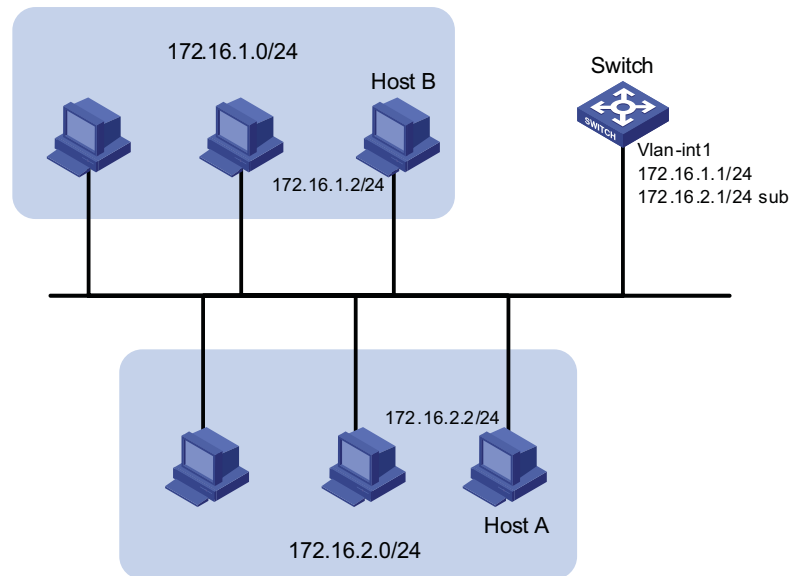
As shown in Figure 37, VLAN-interface 1 on Switch is connected to a LAN comprising two segments: 172.16.1.0/24 and 172.16.2.0/24.

To enable the hosts on the two network segments to access the external network through the switch, and enable the hosts on the two network segments to communicate with each other, do the following:

- Assign a primary IP address and a secondary IP address to VLAN-interface 1 on the switch.
- Set the switch as the gateway on all hosts.

Network diagram

Figure 37 Network diagram for IP addressing configuration



Configuration procedure

Assign a primary IP address and a secondary IP address to VLAN-interface 1.

```
<Switch> system-view
[Switch] interface vlan-interface 1
[Switch-Vlan-interface1] ip address 172.16.1.1 255.255.255.0
[Switch-Vlan-interface1] ip address 172.16.2.1 255.255.255.0 sub
```

Set the gateway address to 172.16.1.1 on the PCs attached to the subnet 172.16.1.0/24, and to 172.16.2.1 on the PCs attached to the subnet 172.16.2.0/24.

Use the **ping** command to verify the connectivity between the switch and the hosts on the subnet 172.16.1.0/24.

```
<Switch> ping 172.16.1.2
PING 172.16.1.2: 56 data bytes, press CTRL_C to break
  Reply from 172.16.1.2: bytes=56 Sequence=1 ttl=255 time=25 ms
  Reply from 172.16.1.2: bytes=56 Sequence=2 ttl=255 time=27 ms
  Reply from 172.16.1.2: bytes=56 Sequence=3 ttl=255 time=26 ms
  Reply from 172.16.1.2: bytes=56 Sequence=4 ttl=255 time=26 ms
  Reply from 172.16.1.2: bytes=56 Sequence=5 ttl=255 time=26 ms

--- 172.16.1.2 ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 25/26/27 ms
```

The information shown above indicates the switch can communicate with the hosts on the subnet 172.16.1.0/24.

Use the **ping** command to verify the connectivity between the switch and the hosts on the subnet 172.16.2.0/24.

```

<Switch> ping 172.16.2.2
PING 172.16.2.2: 56 data bytes, press CTRL_C to break
  Reply from 172.16.2.2: bytes=56 Sequence=1 ttl=255 time=25 ms
  Reply from 172.16.2.2: bytes=56 Sequence=2 ttl=255 time=26 ms
  Reply from 172.16.2.2: bytes=56 Sequence=3 ttl=255 time=26 ms
  Reply from 172.16.2.2: bytes=56 Sequence=4 ttl=255 time=26 ms
  Reply from 172.16.2.2: bytes=56 Sequence=5 ttl=255 time=26 ms

--- 172.16.2.2 ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 25/25/26 ms

```

The information shown above indicates the switch can communicate with the hosts on the subnet 172.16.2.0/24.

Use the **ping** command to verify the connectivity between hosts on the subnet 172.16.1.0/24 and hosts on subnet 172.16.2.0/24. Ping Host B on Host A to verify that the ping operation is successful.

Displaying and Maintaining IP Addressing

To do...	Use the command...	Remarks
Display information about a specified or all Layer 3 interfaces	display ip interface [<i>interface-type interface-number</i>]	Available in any view
Display brief information about a specified or all Layer 3 interfaces	display ip interface brief [<i>interface-type interface-number</i>]	

13

IP PERFORMANCE CONFIGURATION

When configuring IP performance, go to these sections for information you are interested in:

- "IP Performance Overview" on page 127
- "Enabling Reception and Forwarding of Directed Broadcasts to a Directly Connected Network" on page 127
- "Configuring TCP Attributes" on page 129
- "Configuring ICMP to Send Error Packets" on page 130
- "Displaying and Maintaining IP Performance" on page 132

IP Performance Overview

In some network environments, you need to adjust the IP parameters to achieve best network performance. IP performance configuration includes:

- Enabling the device to receive and forward directed broadcasts
- Configuring the maximum TCP segment size (MSS) of the interface
- Configuring TCP timers
- Configuring the TCP buffer size
- Enabling ICMP error packets sending

Enabling Reception and Forwarding of Directed Broadcasts to a Directly Connected Network

Directed broadcasts refer to broadcast packets sent to a specific network. In the destination IP address of a directed broadcast, the network ID is a network-specific number and the host ID is all ones. Enabling the device to receive and forward directed broadcasts to a directly connected network will give hackers an opportunity to attack the network. Therefore, the device is disabled from receiving and forwarding directed broadcasts by default. You should however enable the feature when:

- Using the UDP Helper function to convert broadcasts to unicasts and forward them to a specified server.
- Using the Wake on LAN function to forward directed broadcasts to a PC on the remote network.

Enabling Reception of Directed Broadcasts to a Directly Connected Network

If a device is enabled to receive directed broadcasts, the device will determine whether to forward them according to the configuration on the outgoing interface.

Follow these steps to enable the device to receive directed broadcasts:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the device to receive directed broadcasts	ip forward-broadcast	Required By default, the device is disabled from receiving directed broadcasts.

Enabling Forwarding of Directed Broadcasts to a Directly Connected Network

Follow these steps to enable the device to forward directed broadcasts:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable the interface to forward directed broadcasts	ip forward-broadcast [acl <i>acl-number</i>]	Required By default, the device is disabled from forwarding directed broadcasts.



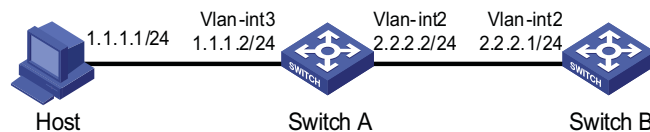
- You can reference an ACL to forward only directed broadcasts permitted by the ACL.
- If you execute the **ip forward-broadcast acl** command on an interface repeatedly, the last execution overwrites the previous one. If the command executed last time does not include the **acl** *acl-number*, the ACL configured previously will be removed.

Configuration Example Network requirements

As shown in Figure 38, the host's interface and VLAN-interface 3 of Switch A are on the same network segment (1.1.1.0/24). VLAN-interface 2 of Switch A and VLAN-interface 2 of Switch B are on another network segment (2.2.2.0/24). The default gateway of the host is VLAN-interface 3 (IP address 1.1.1.2/24) of Switch A. Configure a static route on Switch B to enable the reachability between host and Switch B.

Network diagram

Figure 38 Network diagram for receiving and forwarding directed broadcasts



Configuration procedure

- Configure Switch A

Enable Switch A to receive directed broadcasts.

```
<SwitchA> system-view
[SwitchA] ip forward-broadcast
```

Configure IP addresses for VLAN-interface 3 and VLAN-interface 2.

```
[SwitchA] interface vlan-interface 3
[SwitchA-Vlan-interface3] ip address 1.1.1.2 24
[SwitchA-Vlan-interface3] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 2.2.2.2 24

# Enable VLAN-interface 2 to forward directed broadcasts.

[SwitchA-Vlan-interface2] ip forward-broadcast
```

- Configure Switch B

```
# Enable Switch B to receive directed broadcasts.

<SwitchB> system-view
[SwitchB] ip forward-broadcast

# Configure a static route to the host.

[SwitchB] ip route-static 1.1.1.1 24 2.2.2.2

# Configure an IP address for VLAN-interface 2.

[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 2.2.2.1 24
```

After the above configurations, if you ping the subnet broadcast address (2.2.2.255) of VLAN-interface 2 of Switch A on the host, the ping packets can be received by VLAN-interface 2 of Switch B. However, if you disable the **ip forward-broadcast** command, the ping packets can not be received by the VLAN-interface 2 of Switch B.

Configuring TCP Attributes

Configuring TCP Optional Parameters

TCP optional parameters that can be configured include:

- **synwait** timer: When sending a SYN packet, TCP starts the synwait timer. If no response packets are received within the synwait timer timeout, the TCP connection is not successfully created.
- **finwait** timer: When the TCP connection is in FIN_WAIT_2 state, finwait timer will be started. If no FIN packets are received within the timer timeout, the TCP connection will be terminated. If FIN packets are received, the TCP connection state changes to TIME_WAIT. If non-FIN packets are received, the system restarts the timer from receiving the last non-FIN packet. The connection is broken after the timer expires.
- Size of TCP receive/send buffer

Follow these steps to configure TCP optional parameters:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Configure TCP synwait timer's timeout value	tcp timer syn-timeout <i>time-value</i>	Optional By default, the timeout value is 75 seconds.
Configure TCP finwait timer's timeout value	tcp timer fin-timeout <i>time-value</i>	Optional By default, the timeout value is 675 seconds.
Configure the size of TCP receive/send buffer	tcp window <i>window-size</i>	Optional By default, the buffer is 8 kilobytes.



CAUTION: The actual length of the finwait timer is determined by the following formula:

Actual length of the finwait timer = (Configured length of the finwait timer - 75) + configured length of the synwait timer

Configuring ICMP to Send Error Packets

Sending error packets is a major function of ICMP protocol. In case of network abnormalities, ICMP packets are usually sent by the network or transport layer protocols to notify corresponding devices so as to facilitate control and management.

Advantage of sending ICMP error packets

There are three kinds of ICMP error packets: redirect packets, timeout packets and destination unreachable packets. Their sending conditions and functions are as follows.

1 Sending ICMP redirect packets

A host may have only a default route to the default gateway in its routing table after startup. The default gateway will send ICMP redirect packets to the source host and notify it to reselect a correct next hop router to send the subsequent packets, if the following conditions are satisfied:

- The receiving and forwarding interfaces are the same.
- The selected route has not been created or modified by ICMP redirect packet.
- The selected route is not the default route of the device.
- There is no source route option in the packet.

ICMP redirect packets function simplifies host administration and enables a host to gradually establish a sound routing table to find out the best route

2 Sending ICMP timeout packets

If the device received an IP packet with a timeout error, it drops the packet and sends an ICMP timeout packet to the source.

The device will send an ICMP timeout packet under the following conditions:

- If the device finds the destination of a packet is not itself and the TTL field of the packet is 1, it will send a "TTL timeout" ICMP error message.

- When the device receives the first fragment of an IP datagram whose destination is the device itself, it will start a timer. If the timer times out before all the fragments of the datagram are received, the device will send a "reassembly timeout" ICMP error packet.
- Sending ICMP destination unreachable packets

If the device receives an IP packet with the destination unreachable, it will drop the packet and send an ICMP destination unreachable error packet to the source.

Conditions for sending this ICMP packet:

- If neither a route nor the default route for forwarding a packet is available, the device will send a "network unreachable" ICMP error packet.
- If the destination of a packet is local while the transport layer protocol of the packet is not supported by the local device, the device sends a "protocol unreachable" ICMP error packet to the source.
- When receiving a packet with the destination being local and transport layer protocol being UDP, if the packet's port number does not match the running process, the device will send the source a "port unreachable" ICMP error packet.
- If the source uses "strict source routing" to send packets, but the intermediate device finds the next hop specified by the source is not directly connected, the device will send the source a "source routing failure" ICMP error packet.
- When forwarding a packet, if the MTU of the sending interface is smaller than the packet but the packet has been set "Don't Fragment", the device will send the source a "fragmentation needed and Don't Fragment (DF)-set" ICMP error packet.

Disadvantage of sending ICMP error packets

Although sending ICMP error packets facilitate network control and management, it still has the following disadvantages:

- Sending a lot of ICMP packets will increase network traffic.
- If receiving a lot of malicious packets that cause it to send ICMP error packets, the device's performance will be reduced.
- As the redirection function increases the routing table size of a host, the host's performance will be reduced if its routing table becomes very large.
- If a host sends malicious ICMP destination unreachable packets, end users may be affected.

To prevent such problems, you can disable the device from sending ICMP error packets.

Follow these steps to disable sending ICMP error packets:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Disable sending ICMP redirection packets	undo ip redirects	Required Enabled by default.

To do...	Use the command...	Remarks
Disable sending ICMP timeout packets	undo ip ttl-expires	Required Enabled by default.
Disable sending ICMP destination unreachable packets	undo ip unreachable	Required Enabled by default.



- *The device stops sending "network unreachable" and "source route failure" ICMP error packets after sending ICMP destination unreachable packets is disabled. However, other destination unreachable packets can be sent normally.*
- *The device stops sending "TTL timeout" ICMP error packets after sending ICMP timeout packets is disabled. However, "reassembly timeout" error packets will be sent normally.*

Displaying and Maintaining IP Performance

To do...	Use the command...	Remarks
Display current TCP connection state	display tcp status	Available in any view
Display TCP connection statistics	display tcp statistics	
Display UDP statistics	display udp statistics	
Display IP packets statistics	display ip statistics	
Display ICMP flows statistics	display icmp statistics	
Display socket information	display ip socket [socketype <i>sock-type</i>] [<i>task-id socket-id</i>]	
Display FIB forward information	display fib [{ begin include exclude } <i>string</i> acl <i>acl-number</i> ip-prefix <i>ip-prefix-name</i>]	
Display FIB forward information matching the specified destination IP address	display fib <i>ip-address1</i> [{ <i>mask1</i> <i>mask-length1</i> } [<i>ip-address2</i> { <i>mask2</i> <i>mask-length2</i> }] longer] longer]	
Display statistics about the FIB items	display fib statistics	
Clear statistics of IP packets	reset ip statistics	Available in user view
Clear statistics of TCP connections	reset tcp statistics	
Clear statistics of UDP flows	reset udp statistics	

14

QINQ CONFIGURATION

When configuring QinQ, go to these sections for information you are interested in:

- "Introduction to QinQ" on page 133
- "Configuring Basic QinQ" on page 135
- "Configuring Selective QinQ" on page 136
- "Configuring the TPID Value to Be Carried in VLAN Tags" on page 137
- "QinQ Configuration Example" on page 137

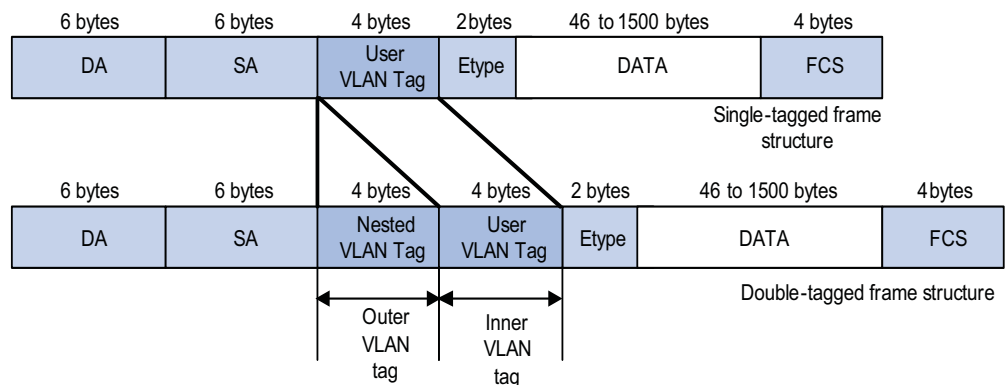
Introduction to QinQ

Understanding QinQ

In the VLAN tag field defined in IEEE 802.1Q, only 12 bits are used for VLAN IDs, so a switch can support a maximum of 4,094 VLANs. In actual applications, however, a large number of VLANs are required to isolate users, especially in metropolitan area networks (MANs), and 4,094 VLANs are far from satisfying such requirements.

The port QinQ feature is a flexible, easy-to-implement Layer 2 VPN technique, which enables the access point to encapsulate an outer VLAN tag in Ethernet frames from customer networks (private networks), so that the Ethernet frames will travel across the service provider's backbone network (public network) with double VLAN tags. The inner VLAN tag is the customer network VLAN tag while the outer one is the VLAN tag assigned by the service provider to the customer. In the public network, frames are forwarded based on the outer VLAN tag only, with the source MAC address learned as a MAC address table entry for the VLAN indicated by the outer tag, while the customer network VLAN tag is transmitted as part of the data in the frames.

Figure 39 shows the structure of 802.1Q-tagged and double-tagged Ethernet frames. The QinQ feature enables a device to support up to 4,094 x 4,094 VLANs to satisfy the requirement for the amount of VLANs in the MAN.

Figure 39 Single-tagged frame structure vs. double-tagged Ethernet frame structure

Advantages of QinQ:

- Addresses the shortage of public VLAN ID resource.
- Enables customers to plan their own VLAN IDs, without running into conflicts with public network VLAN IDs.
- Provides an easy-to-do Layer 2 VPN solution for small-sized MANs or intranets.



The QinQ feature requires configurations only on the service provider network, and not on the customer network.

Implementations of QinQ

There are two types of QinQ implementations: basic QinQ and selective QinQ.

1 Basic QinQ

- Basic QinQ is a port-based feature, which is implemented through VLAN VPN.
- With the VLAN VPN feature enabled on a port, when a frame arrives at the port, the switch will tag it with the port's default VLAN tag, regardless of whether the frame is tagged or untagged. If the received frame is already tagged, this frame becomes a double-tagged frame; if it is an untagged frame, it is tagged with the port's default VLAN tag.

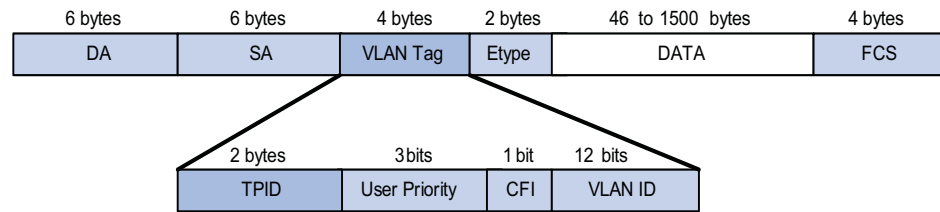
2 Selective QinQ

- Selective QinQ is a more flexible, VLAN-based implementation of QinQ. In addition to all the functions of basic QinQ, selective QinQ can tag the frame with different outer VLAN tags based on different inner VLAN IDs.

Modification of TPID Value of QinQ Frames

A VLAN tag uses the tag protocol identifier (TPID) field to identify the protocol type of the tag. The value of this field, as defined in IEEE 802.1Q, is 0x8100.

Figure 40 shows the 802.1Q-defined tag structure of an Ethernet frame.

Figure 40 VLAN Tag structure of an Ethernet frame

The device determines whether a received frame carries a service provider VLAN tag or a customer VLAN tag by checking the corresponding TPID value. Upon receiving a frame, the device compares the configured TPID value with the value of the TPID field in the frame. If the two match, the frame carries the corresponding VLAN tag. For example, if a frame carries VLAN tags with the TPID values of 0x9100 and 0x8100 respectively while the configured TPID value of the service provider VLAN tag is 0x9100 and that of the VLAN tag for a customer network is 0x8200, the device considers that the frame carries only the service provider VLAN tag but not the customer VLAN tag.

In addition, the systems of different vendors may set the TPID of the outer VLAN tag of QinQ frames to different values. For compatibility with these systems, you can modify the TPID value so that the QinQ frames, when sent to the public network, carry the TPID value identical to the value of a particular vendor to allow interoperability with the devices of that vendor.

The TPID in an Ethernet frame has the same position with the protocol type field in a frame without a VLAN tag. To avoid problems in packet forwarding and handling in the network, you cannot set the TPID value to any of the values in the table below.

Table 28 Reserved protocol type values

Protocol type	Value
ARP	0x0806
PUP	0x0200
RARP	0x8035
IP	0x0800
IPv6	0x86DD
PPPoE	0x8863/0x8864
MPLS	0x8847/0x8848
IPX/SPX	0x8137
IS-IS	0x8000
LACP	0x8809
802.1x	0x888E
Cluster	0x88A7
Reserved	0xFFFFD/0xFFFFE/0xFFFF

Configuring Basic QinQ

Follow these steps to configure basic QinQ:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view or port group view	Enter Ethernet port view Enter port group view	Required Use either command. Configurations made in Ethernet port view will take effect on the current port only; configuration made in port group view will take effect on all ports in the port group.
Enable QinQ on the port(s)	qinq enable	Required Disabled by default.

Configuring Selective QinQ

The outer VLAN tag added to a frame by the basic QinQ feature is the VLAN tag corresponding to the port's default VLAN ID, while the selective QinQ feature allows adding different outer VLAN tags based on different inner VLAN tags.

With selective QinQ configured on a port, the device attaches different outer VLAN tags based on the inner VLAN tags; frames with a VLAN ID out of the range specified in the **raw-vlan-id inbound** command are attached the port's default VLAN tag as the outer tag.

Follow these steps to configure selective QinQ:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view or port group view	Enter Ethernet port view Enter port group view	Required Use either command. Configurations made in Ethernet port view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
Enter QinQ view and configure the outer VLAN tag for the port to add	qinq vid <i>vlan-id</i>	Required
Configure inner VLAN tags corresponding to the outer VLAN tags	raw-vlan-id inbound { all <i>vlan-id-list</i> }	Required



CAUTION:

- An inner VLAN tag corresponds to only one outer VLAN tag. If you want to change an outer VLAN tag, you must delete the old outer VLAN tag configuration and configure a new outer VLAN tag.
- You can configure selective QinQ and basic QinQ on the same port. The switch uses the basic QinQ function to attach the port's default VLAN tag as the outer tag to frames that do not match the selective QinQ mapping rule.

Configuring the TPID Value to Be Carried in VLAN Tags

You can configure the TPID value to be carried in a VLAN tag TPID globally (configuration will take effect on all ports of the device).

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the TPID value to be carried in the customer VLAN tag or the service provider VLAN tag	qinq ethernet-type [customer-tag service-tag] hex-value	Optional Both 0x8100 by default

QinQ Configuration Example

Network requirements

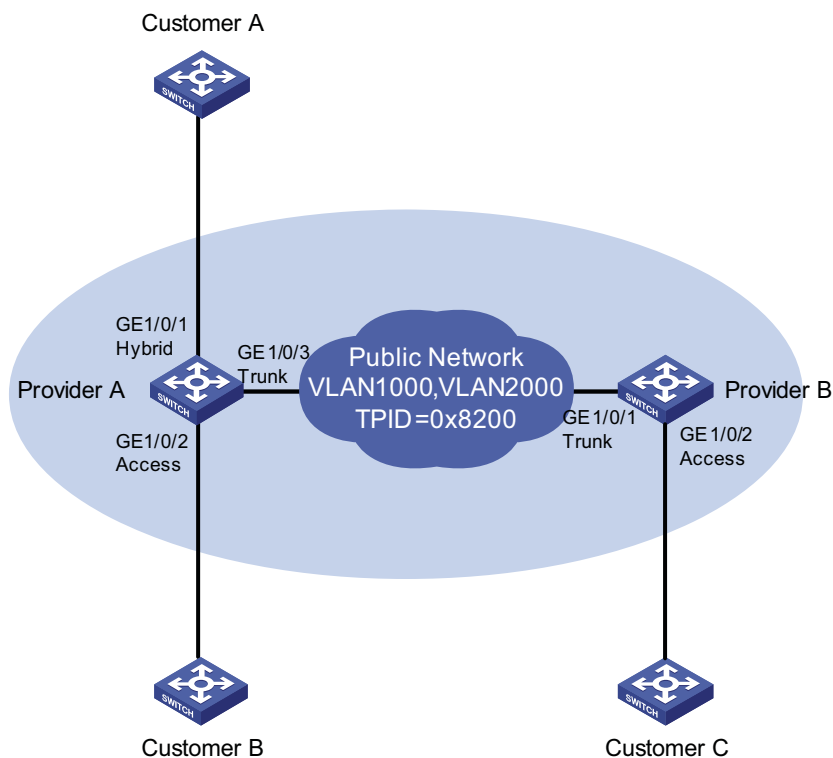
- Provider A and Provider B are service provider network access devices.
- Customer A, Customer B and Customer C are customer network access devices.
- Provider A and Provider B are interconnected through a configured trunk port. Provider A belongs to VLAN 1000 of the service provider network, and Provider B belongs to VLAN 2000 of the service provider network.
- Third-party devices are deployed between Provider A and Provider B, with a TPID value of 0x8200.

After configuration, the network should satisfy the following requirement:

- Frames of VLAN 10 of Customer A and frames of VLAN 10 of Customer B can be forwarded to each other through VLAN 1000 of the provider network; frames of VLAN 20 of Customer A and frames of VLAN 20 of Customer C can be forwarded to each other through VLAN 2000 of the provider network.

Network diagram

Figure 41 Network diagram for QinQ configuration



Configuration procedure



With this configuration, the user must allow the QinQ packets to pass between the devices of the service providers.

1 Configuration on Provider A

Enter system view.

```
<ProviderA> system-view
```

■ Configuration on GigabitEthernet 1/0/1

Configure GigabitEthernet 1/0/1 as a Hybrid port that permits frames of VLAN 1000 and VLAN 2000 to pass, and configure the port to remove the outer tag of the frames when sending them out.

```
[ProviderA] interface GigabitEthernet 1/0/1
[ProviderA-GigabitEthernet1/0/1] port link-type hybrid
[ProviderA-GigabitEthernet1/0/1] port hybrid vlan 1000 2000 untagged
```

Configure the port to tag frames from VLAN 10 with an outer tag with the VLAN ID of 1000.

```
[ProviderA-GigabitEthernet1/0/1] qinq vid 1000
[ProviderA-GigabitEthernet1/0/1-vid-1000] raw-vlan-id inbound 10
[ProviderA-GigabitEthernet1/0/1-vid-1000] quit
```

Configure the port to tag frames from VLAN 20 with an outer tag with the VLAN ID of 2000.

```
[ProviderA-GigabitEthernet1/0/1] qinq vid 2000
[ProviderA-GigabitEthernet1/0/1-vid-2000] raw-vlan-id inbound 20
[ProviderA-GigabitEthernet1/0/1-vid-2000] quit
[ProviderA-GigabitEthernet1/0/1] quit
```

- Configuration on GigabitEthernet 1/0/2

Configure VLAN 1000 as the default VLAN of the port.

```
[ProviderA] interface GigabitEthernet 1/0/2
[ProviderA-GigabitEthernet1/0/2] port access vlan 1000
```

Enable basic QinQ so that the port tags frames from VLAN 10 with an outer tag with the VLAN ID of 1000.

```
[ProviderA-GigabitEthernet1/0/2] qinq enable
[ProviderA-GigabitEthernet1/0/2] quit
```

- Configuration on GigabitEthernet 1/0/3.

Configure GigabitEthernet 1/0/3 as a trunk port, and permit frames of VLAN 1000 and VLAN 2000 to pass.

```
[ProviderA] interface GigabitEthernet 1/0/3
[ProviderA-GigabitEthernet1/0/3] port link-type trunk
[ProviderA-GigabitEthernet1/0/3] port trunk permit vlan 1000 2000
```

To enable interoperability with the third-party devices in the public network, set the TPID value to be carried in VLAN Tags to 0x8200.

```
[ProviderA-GigabitEthernet1/0/3] quit
[ProviderA] qinq ethernet-type service-tag 8200
```

2 Configuration on Provider B

- Configuration on GigabitEthernet 1/0/1

Configure GigabitEthernet 1/0/1 as a trunk port, and permit frames of VLAN 1000 and VLAN 2000.

```
<ProviderB> system-view
[ProviderB] interface GigabitEthernet 1/0/1
[ProviderB-GigabitEthernet1/0/1] port link-type trunk
[ProviderB-GigabitEthernet1/0/1] port trunk permit vlan 1000 2000
```

To enable interoperability with the third-party devices in the public network, set the TPID value to be carried in VLAN Tags to 0x8200.

```
[ProviderB-GigabitEthernet1/0/1] quit
[ProviderB] qinq ethernet-type service-tag 8200
```

- Configuration on GigabitEthernet 1/0/2

Configure VLAN 2000 as the default VLAN of the port.

```
[ProviderB] interface GigabitEthernet 1/0/2  
[ProviderB-GigabitEthernet1/0/2] port access vlan 2000
```

Enable basic QinQ so as to tag frames from VLAN 20 with an outer tag with the VLAN ID of 2000.

```
[ProviderB-GigabitEthernet1/0/2] qinq enable
```

3 Configuration on devices on the public network

As third-party devices are deployed between Provider A and Provider B, what we discuss here is only the basic configuration that should be made on the devices. Configure that device connecting with GigabitEthernet 1/0/3 of Provider A and the device connecting with GigabitEthernet 1/0/1 of Provider B so that their corresponding ports send tagged frames of VLAN 1000 and VLAN 2000. The configuration steps are omitted here.

15

BPDU TUNNELING CONFIGURATION

When configuring BPDU tunneling, go to these sections for information you are interested in:

- "Introduction to BPDU Tunneling" on page 141
- "Configuring BPDU Isolation" on page 142
- "Configuring BPDU Transparent Transmission" on page 143
- "Configuring Destination Multicast MAC Address for BPDU Tunnel Frames" on page 144
- "BPDU Tunneling Configuration Example" on page 144

Introduction to BPDU Tunneling

Why BPDU Tunneling

To avoid loops in your network, you can enable the spanning tree protocol (STP) on your device. However, STP gets aware of the topological structure of a network by means of bridge protocol data units (BPDUs) exchanged between different devices and the BPDUs are Layer 2 multicast packets, which can be received and processed by all STP-enabled devices on the network. This prevents each network from correctly calculating its spanning tree. As a result, when redundant links exist in a network, data loops will unavoidably occur.

By allowing each network to have its own spanning tree while running STP, BPDU tunneling can resolve this problem.

- BPDU tunneling can isolate BPDUs of different customer networks, so that one network is not affected by others while calculating the topological structure.
- BPDU tunneling enables BPDUs of the same customer network to be broadcast in a specific VLAN in the provider network, so that the geographically dispersed customer networks of the same customer can implement consistent spanning tree calculation across the provider network.

How BPDU Tunneling Works

The BPDU tunneling implements the following two functions:

- BPDU isolation
- BPDU transparent transmission

The work process of IGMP is as follows:

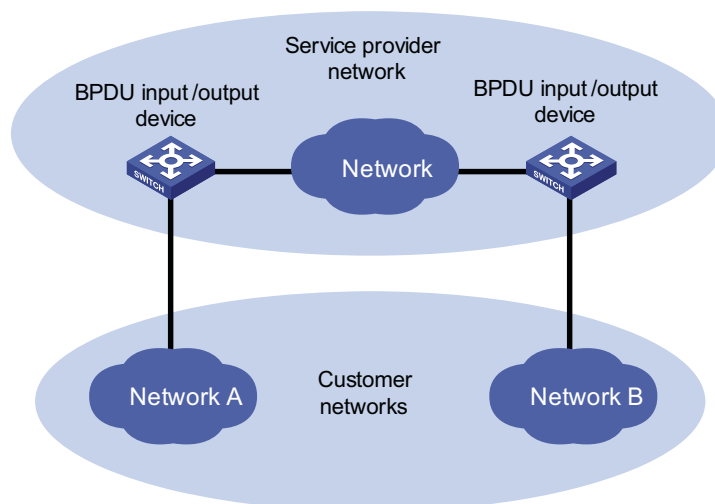
BPDU isolation

When a port receives BPDUs of other networks, the port will discard the BPDUs, so that they will not take part in spanning tree calculation. Refer to “Configuring BPDU Isolation” on page 142.

BPDU transparent transmission

As shown in Figure 42, the upper part is the service provider network, and the lower part represents the customer networks. The customer networks include network A and network B. Enabling the BPDU tunneling function on the BPDU input/output devices across the service provider network allows BPDUs of the customer networks to be transparently transmitted in the service provider network, and allows each customer network to implement independent spanning tree calculation, without affecting each other. Refer to “Configuring BPDU Transparent Transmission” on page 143.

Figure 42 Network hierarchy of BPDU tunneling



- At the BPDU input side, the device changes the destination MAC address of a BPDU from a customer network from 0x0180-C200-0000 to a special multicast MAC address, 0x010F-E200-0003 by default. In the service provider’s network, the modified BPDUs are forwarded as data packets in the user VLAN.
- At the packet output side, the device recognizes the BPDU with the destination MAC address of 0x010F-E200-0003 and restores its original destination MAC address 0x0180-C200-0000. Then, the device removes the outer tag, and sends the BPDU to the destination customer network.



Make sure, through configuration, that the VLAN tag of the BPDU is neither changed nor removed during its transparent transmission in the service provider network; otherwise, the system will fail to transparently transmit the customer network BPDU correctly.

Configuring BPDU Isolation

Perform the following tasks to configure BPDU isolation:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Enable BPDU tunneling globally	bpdu-tunnel dot1q enable	Optional Enabled by default
Enter Ethernet port view or port group view	Enter Ethernet port view Enter port group view interface <i>interface-type</i> <i>interface-number</i> port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Required Use either command. Configurations made in Ethernet port view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
Enable BPDU tunneling for the port(s)	bpdu-tunnel dot1q enable	Required Disabled by default



- *BPDU tunneling must be enabled globally before the BPDU tunnel configuration for a port can take effect.*
- *The BPDU tunneling feature is incompatible with the GVRP feature, so these two features cannot be enabled at the same time. For introduction to GVRP, refer to “Introduction to GVRP” on page 109.*
- *The BPDU tunneling feature is incompatible with the NTDP feature, so these two features cannot be enabled at the same time. If you want to enable BPDU tunneling on a port, use the **undo ntdp enable** command to disable NTDP first. For introduction to NTDP, refer to “Cluster Management Overview” on page 905.*

Configuring BPDU Transparent Transmission

Perform the following tasks to configure BPDU transparent transmission:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable BPDU tunneling globally	bpdu-tunnel dot1q enable	Optional Enabled by default
Enter Ethernet port view or port group view	Enter Ethernet port view Enter port group view interface <i>interface-type</i> <i>interface-number</i> port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Required Use either command. Configurations made in Ethernet port view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
Enable BPDU tunneling on the port(s)	bpdu-tunnel dot1q enable	Required Disabled by default
Disable STP on the port(s)	stp disable	Required Enabled by default
Enable BPDU tunneling for STP on the port(s)	bpdu-tunnel dot1q stp	Required Disabled by default



- *BPDU tunneling must be enabled globally before the BPDU tunnel configuration for a port can take effect.*
- *The BPDU tunneling feature is incompatible with the GVRP feature, so these two features cannot be enabled at the same time. For introduction to GVRP, refer to “Introduction to GVRP” on page 109.*
- *The BPDU tunneling feature is incompatible with the NTDP feature, so these two features cannot be enabled at the same time. If you want to enable BPDU tunneling on a port, use the **undo ntdp enable** command to disable NTDP first. For introduction to NTDP, refer to “Cluster Management Overview” on page 905.*

Configuring Destination Multicast MAC Address for BPDU Tunnel Frames

By default, the destination multicast MAC address for BPDU Tunnel frames is 0x010F-E200-0003. You can modify it to 0x0100-0CCD-CDD0, 0x0100-0CCD-CDD1 or 0x0100-0CCD-CDD2 through the following configuration.

Follow these steps to configure destination multicast MAC address for BPDU tunnel frames:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the destination multicast MAC address for BPDU Tunnel frames	bpdu-tunnel tunnel-dmac <i>mac-address</i>	Optional 0x010F-E200-0003 by default.

BPDU Tunneling Configuration Example

Network requirements

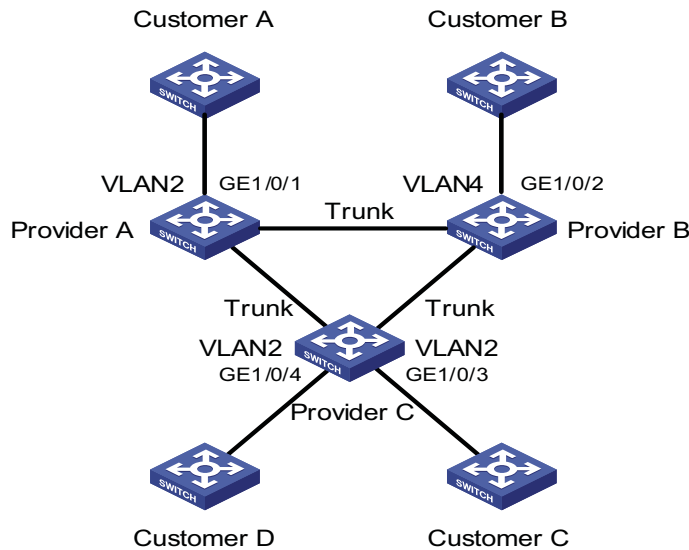
- Customer A, Customer B, Customer C, and Customer D are customer network access devices.
- Provider A, Provider B, and Provider C are service provider network access devices, which are interconnected through configured trunk ports.

The configuration is required to satisfy the following requirements:

- Geographically dispersed customer network devices Customer A, Customer C and Customer D can implement consistent spanning tree calculation across the service provider network.
- BPDU packets from Customer B are isolated so it does not take part in the spanning tree calculation.

Network diagram

Figure 43 Network diagram for BPDU tunneling configuration



Configuration procedure

1 Configuration on Provider A

Configure BPDU transparent transmission on GigabitEthernet 1/0/1.

```
<ProviderA> system-view
[ProviderA] interface GigabitEthernet 1/0/1
[ProviderA-GigabitEthernet1/0/1] port access vlan 2
[ProviderA-GigabitEthernet1/0/1] stp disable
[ProviderA-GigabitEthernet1/0/1] undo ntdp enable
[ProviderA-GigabitEthernet1/0/1] bpdu-tunnel dot1q enable
[ProviderA-GigabitEthernet1/0/1] bpdu-tunnel dot1q stp
```

2 Configuration on Provider B

Configure BPDU isolation on GigabitEthernet 1/0/2.

```
<ProviderB> system-view
[ProviderB] interface GigabitEthernet 1/0/2
[ProviderB-GigabitEthernet1/0/2] port access vlan 4
[ProviderB-GigabitEthernet1/0/2] undo ntdp enable
[ProviderB-GigabitEthernet1/0/2] bpdu-tunnel dot1q enable
```

3 Configuration on Provider C

Configure BPDU transparent transmission on GigabitEthernet 1/0/3.

```
<ProviderC> system-view
[ProviderC] interface GigabitEthernet 1/0/3
[ProviderC-GigabitEthernet1/0/3] port access vlan 2
[ProviderC-GigabitEthernet1/0/3] stp disable
[ProviderC-GigabitEthernet1/0/3] undo ntdp enable
[ProviderC-GigabitEthernet1/0/3] bpdu-tunnel dot1q enable
[ProviderC-GigabitEthernet1/0/3] bpdu-tunnel dot1q stp
```

Configure BPDU transparent transmission on GigabitEthernet 1/0/4.

```
[ProviderC-GigabitEthernet1/0/3] quit
[ProviderC] interface GigabitEthernet 1/0/4
[ProviderC-GigabitEthernet1/0/4] port access vlan 2
[ProviderC-GigabitEthernet1/0/4] stp disable
[ProviderC-GigabitEthernet1/0/4] undo ntdp enable
[ProviderC-GigabitEthernet1/0/4] bpdu-tunnel dot1q enable
[ProviderC-GigabitEthernet1/0/4] bpdu-tunnel dot1q stp
```



When STP works stably on the customer network, if Customer A acts as the root bridge, the ports of Customer C and Customer D connected with Provider C can receive BPDUs from Customer A. Since BPDU isolation is enabled on Customer B, the port that connects Customer B to Provider B cannot receive BPDUs from Customer A.

16

PORT CORRELATION CONFIGURATION

When configuring Ethernet ports, go to these sections for information you are interested in:

- "Ethernet Port Configuration" on page 147
- "Maintaining and Displaying an Ethernet Port" on page 156

Ethernet Port Configuration

Complete the following tasks to configure an Ethernet port:

Task	Remarks
"Performing Basic Ethernet Port Configuration" on page 147	Optional
"Combo Port Configuration" on page 148	Optional
"Configuring the Suppression Time of Physical-Link-State Change on an Ethernet Port" on page 149	Optional
"Enabling Loopback Test on an Ethernet Port" on page 149	Optional
"Configuring a Port Group" on page 150	Optional
"Configuring the Broadcast/Multicast/Unknown Unicast Storm Suppression Ratio for an Ethernet Port" on page 151	Optional
"Setting the Interval for Collecting Ethernet Port Statistics" on page 152	Optional
"Enabling Forwarding of Jumbo Frames" on page 152	Optional
"Enabling Loopback Detection on an Ethernet Port" on page 153	Optional
"Configuring the Cable Type for an Ethernet Port" on page 153	Optional
"Testing the Cable on an Ethernet Port" on page 154	Optional
"Configuring the Storm Constrain Function on an Ethernet Port" on page 155	Optional

Performing Basic Ethernet Port Configuration

Three types of duplex modes are available to Ethernet ports:

- Full-duplex mode (full). Ports operating in this mode can send and receive packets simultaneously.
- Half-duplex mode (half). Ports operating in this mode can either send or receive packets at a given time.
- Auto-negotiation mode (auto). Ports operating in this mode determine their duplex mode through auto-negotiation.

Similarly, if you configure the transmission rate for an Ethernet port by using the **speed** command with the **auto** keyword specified, the transmission rate is determined through auto-negotiation too.

Follow these steps to perform basic Ethernet port configuration:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	-
Set the description string	description <i>text</i>	Optional By default, the description string is "interface index + Interface".
Set the duplex mode	duplex { auto full half }	Optional auto by default.
Set the transmission rate	speed { 10 100 1000 auto }	Optional auto by default.
Shut down the Ethernet port	shutdown	Optional By default, an Ethernet port is in up state. To bring up an Ethernet port, use the undo shutdown command.



The **speed 1000** command is only applicable to GigabitEthernet ports.

Combo Port Configuration

Introduction to Combo port

A Combo port can operate as either an optical port or an electrical port. Inside the device there is only one forwarding interface. For a Combo port, the electrical port and the corresponding optical port are TX-SFP multiplexed. You can specify a Combo port to operate as an electrical port or an optical port. That is, a Combo port cannot operate as both an electrical port and an optical port simultaneously.

For ease of management, a Combo port can be categorized into one of the following two types:

- Single Combo port: the two Ethernet interfaces in the device panel correspond to only one interface view, in which state on the two interfaces can be realized. A single Combo port can be a Layer 2 Ethernet interface or a Layer 3 Ethernet interface.
- Dual-Combo port: the two Ethernet interfaces in the device panel correspond to two interface views. State switchover can be realized in user's own interfaces view. A double Combo port can only be a layer 2 Ethernet interface.



Currently, only Dual-Combo ports are supported on the Switch 4800G.

Configuring Combo port state

Follow these steps to configure the state for a double Combo port:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet interface view	interface <i>interface-type</i> <i>interface-number</i>	-

To do...	Use the command...	Remarks
Enable a specified double Combo port	undo shutdown	Optional By default, out of the two ports in a Combo port, the one with a smaller port ID is enabled.

For detailed information about Combo ports and the corresponding physical ports, refer to the installation manual.

Enabling Flow Control on an Ethernet Port

When flow control is enabled on both sides, if traffic congestion occurs on one side, the side will send a Pause frame notifying the peer side to temporarily suspend the sending of packets. The peer side is expected to stop sending packets when it receives the Pause frame. In this way, flow controls helps to avoid the dropping of packets. Note that flow control can take effect only when it is enabled on both sides.

Follow these steps to enable flow control on an Ethernet port:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view	interface <i>interface-type interface-number</i>	-
Enable flow control	flow-control	Required Turned off by default

Configuring the Suppression Time of Physical-Link-State Change on an Ethernet Port

An Ethernet port operates in one of the two physical link states: up or down. During the suppression time, physical-link-state changes will not be propagated to the system. Only after the suppression time has elapsed will the system be notified of the physical-link-state changes by the physical layer. This functionality reduces the extra overhead occurred due to frequent physical-link-state changes within a short period of time.

Follow these steps to configure the suppression time of physical-link-state changes on an Ethernet port:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view	interface <i>interface-type interface-number</i>	-
Configure the up/down suppression time of physical-link-state changes	link-delay <i>delay-time</i>	Required The default suppression time is 0 seconds, indicating that the physical layer reports the change of the port state to the system right after the port state changes.

Enabling Loopback Test on an Ethernet Port

You can enable loopback testing to check whether the Ethernet port functions properly. Note that no data packets can be forwarded during the testing. Loopback testing falls into the following two categories:

- Internal loopback test, which is performed within switching chips to test the functions related to the Ethernet ports.
- External loopback test, which is used to test the hardware functions of an Ethernet port. To perform external loopback testing on an Ethernet port, you need to install a loopback plug on the Ethernet port. In this case, packets sent from the port are received by the same port.

Follow these steps to enable Ethernet port loopback test:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable loopback test	loopback { external internal }	Optional Disabled by default.



- *As for the internal loopback test and external loopback test, if a port is down, only the former is available on it; if the port is shut down, both are unavailable.*
- *The **speed**, **duplex**, **mdi**, and **shutdown** commands are not applicable during a loopback test.*
- *With the loopback test enabled, the Ethernet port operates in the full duplex mode. With the loopback test enabled, the original configurations will be restored.*

Configuring a Port Group

To make the configuration task easier for users, certain devices allow users to configure on a single port as well as on multiple ports in a port group. In port group view, the user only needs to input the configuration command once on one port and that configuration will apply to all ports in the port group. This effectively reduces redundant configurations.

A Port group belongs to one of the following two categories:

- Manual port group: manually created by users. Multiple Ethernet ports can be added to the same port group;
- Dynamic port group: dynamically created by the system. Currently, it refers in particular to a port aggregation group. A port aggregation port group is automatically created together with the creation of a link aggregation group and cannot be created by users through CLI. Adding or deleting of ports in a port aggregation port group can only be achieved through operations on the link aggregation group.

A port group enables you to configure ports in batch. You cannot display or save the configuration of a port group. However, you can use the **display current-configuration** or **display this** command to view the current configuration of each member port of a port group.

Follow these steps to configure a port group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter port group view	Enter manual port group view port-group manual <i>port-group-name</i>	-
	Enter aggregation port group view port-group aggregation <i>agg-id</i>	-

Follow these steps to configure manual port group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a manual port group and enter manual port group view	port-group manual <i>port-group-name</i>	Required
Add Ethernet ports to the manual port group	group-member <i>interface-list</i>	Required



For more information, refer to “Aggregation Port Group” on page 166.

Configuring the Broadcast/Multicast/Unknown Unicast Storm Suppression Ratio for an Ethernet Port

You can use the following commands to suppress the broadcast, multicast, and unknown unicast traffic. In port configuration mode, the suppression ratio indicates the maximum broadcast, multicast, or unknown unicast traffic that is allowed to pass through a port. When the broadcast, multicast, or unknown unicast traffic passing the port exceeds the threshold, the system will discard the extra packets so that the broadcast, multicast, or unknown unicast traffic ratio can drop below the limit to ensure that the network functions properly.



The storm suppression ratio settings configured for an Ethernet port may get invalid if you configure a traffic threshold for the port using the **storm-constrain** command.

Follow these steps to set the broadcast/multicast/unknown unicast storm suppression ratios:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view or port group view	Enter Ethernet port view interface <i>interface-type</i> <i>interface-number</i> Enter port group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Either is required. If configured in Ethernet port view, this feature takes effect on the current port only; if configured in port group view, this feature takes effect on all the ports in the port group.
Configure broadcast storm suppression ratio	broadcast-suppression { <i>ratio</i> pps <i>max-pps</i> }	Optional By default, all broadcast traffic is allowed to pass through a port, that is, broadcast traffic is not suppressed.

To do...	Use the command...	Remarks
Configure multicast storm suppression ratio	multicast-suppression { <i>ratio</i> pps <i>max-pps</i> }	Optional By default, all multicast traffic is allowed to pass through a port, that is, multicast traffic is not suppressed.
Configure unknown unicast storm suppression ratio	unicast-suppression { <i>ratio</i> pps <i>max-pps</i> }	Optional By default, all unknown unicast traffic is allowed to pass through a port, that is, unknown unicast traffic is not suppressed.



If you set storm suppression ratios in Ethernet port view or port group view repeatedly for an Ethernet port that belongs to a port group, only the latest settings take effect.

Setting the Interval for Collecting Ethernet Port Statistics

Follow these steps to configure the interval for collecting port statistics:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the interval for collecting port statistics	interface <i>interface-type</i> <i>interface-number</i> flow-interval <i>interval</i>	Optional By default, the interval for collecting port statistics is 300 seconds.

Enabling Forwarding of Jumbo Frames

Due to tremendous amount of traffic occurring in Ethernet, it is likely that some frames might have a frame size greater than the standard Ethernet frame size. By allowing such frames (called jumbo frames) to pass through Ethernet ports, you can forward frames with a size greater than the standard Ethernet frame size and yet still within the specified parameter range.

You can set the jumbo frame length in Ethernet port view or port group view.

- If you set the jumbo frame length in Ethernet port view, the configuration takes effect only on the current port.
- If you set the jumbo frame length in port group view, the configuration takes effect on all ports in the port group.

Follow these steps to enable the forwarding of jumbo frames:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter the corresponding view	Enter port-group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Use either approach.
Enter Ethernet port view	Enter Ethernet port view interface <i>interface-type</i> <i>interface-number</i>	
Set the maximum frame length allowed on an Ethernet port to 9212 bytes	jumboframe enable	By default, the maximum frame length allowed on an Ethernet port is 9212 bytes.

Enabling Loopback Detection on an Ethernet Port

Loop occurs when a port receives the packets that it sent out. Loops may cause broadcast storm. The purpose of loopback detection is to detect loops on a port.

With loopback detection enabled on an Ethernet port, the device checks the port for external loopback periodically. Once a loopback is detected on the port, the system does the following:

- If loops are detected on a port that is of access type, the port will be shutdown. Meanwhile, trap messages will be sent to the terminal, and the corresponding MAC address forwarding entries will be removed.
- If loops are detected on a port that is of trunk or hybrid type, trap messages are sent to the terminal. If the loopback detection control function is also enabled on the port, the port will be blocked, trap messages will be sent to the terminal, and the corresponding MAC address forwarding entries will be removed.

Follow these steps to configure loopback detection:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable global loopback detection	loopback-detection enable	Required Disabled by default
Configure the interval for port loopback detection	loopback-detection interval-time <i>time</i>	Optional 30 seconds by default
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable loopback detection on the port	loopback-detection enable	Required Disabled by default
Enable loopback detection control on the port (Trunk or Hybrid)	loopback-detection control enable	Optional Disabled by default
Enable loopback detection in all the VLANs containing the port	loopback-detection per-vlan enable	Optional Enabled only in the default VLAN(s) with Trunk port or Hybrid ports



CAUTION:

- Loopback detection on a given port is enabled only after the **loopback-detection enable** command has been issued in both system view and the port view of the port.
- Loopback detection on all ports will be disabled after the issuing of the **undo loopback-detection enable** command in system view.
- If the system detects loopback in multiple VLANs on a port in a detection interval, it sends only one trap to the terminal rather than one trap per VLAN.
- The aggregation port can not support loopback detection.

Configuring the Cable Type for an Ethernet Port

Two types of Ethernet cables can be used to connect Ethernet devices: crossover cable and straight-through cable. To accommodate these two types of cables, an

Ethernet interface on a device can operate in one of the following three Medium Dependent Interface (MDI) modes:

- Across mode, where the Ethernet interface only accepts crossover cables.
- Normal mode, where the Ethernet interface only accepts straight-through cables.
- Auto mode, where the Ethernet interface accepts both straight-through cables and crossover cables.

Normally, the auto mode is recommended. The other two modes are useful only when the device cannot determine the cable type.

Follow these steps to configure the cable type for an Ethernet Port:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure the cable type the Ethernet port can identify	mdi { across auto normal }	Optional Defaults to auto. That is, the Ethernet port automatically detects the type of the cable in use.



*10 GE port cannot support **mdi** configuration.*

Testing the Cable on an Ethernet Port



A link in the up state goes down and then up automatically if you perform the operation described in this section on one of the Ethernet ports forming the link.

You can enable the test on the cable connected with an Ethernet port to check:

- Whether the RX and TX of the cable are short-circuited.
- Whether the cable is open circuited.
- The length of the faulty cable if there is any fault.

The system will return the check result in 5 seconds.

Follow these steps to test the current operating state of the cable connected to an Ethernet port:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	-
Test the current operating state of the cable connected to the port	virtual-cable-test	Required

Configuring the Storm Constrain Function on an Ethernet Port

The storm constrain function suppresses packet storm in an Ethernet. With this function enabled on a port, the system detects the unicast traffic, multicast traffic, or broadcast traffic passing through the port periodically and takes corresponding actions (that is, blocking or shutting down the port and sending trap messages and logs) if the traffic detected exceeds the threshold.



CAUTION: Although the storm suppression function and the storm constrain function can all be used to control specific type of traffic, they conflict with each other. So, do not configure the both for an Ethernet port at the same time. For example, with multicast storm suppression ratio set on an Ethernet port, do not enable the storm constrain function for multicast traffic on the port. Refer to “Configuring the Broadcast/Multicast/Unknown Unicast Storm Suppression Ratio for an Ethernet Port” on page 151 for information about the storm suppression function.

With the storm constrain function enabled on an Ethernet port, you can specify the system to act as follows when the traffic detected exceeds the threshold.

- Blocking the port. In this case, the port is blocked and thus stops forwarding the traffic of this type till the traffic detected is lower than the threshold. Note that a port blocked by the storm constrain function can still forward other types of traffic and monitor the blocked traffic.
- Shutting down the port. In this case, the port is shut down and stops forwarding all types of traffics. Ports shut down by the storm constrain function can only be brought up by using the **undo shutdown** command or disabling the storm constrain function.

Follow these steps to configure the storm constrain function on an Ethernet port:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Set the interval for generating traffic statistics	storm-constrain interval <i>seconds</i>	Optional Defaults to 10 seconds.
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable the storm constrain function and set the lower threshold and the upper threshold	storm-constrain { broadcast multicast } pps <i>max-pps-values</i> <i>min-pps-values</i>	Required By default, the storm constrain function is disabled.
Set the action to be taken when the traffic exceeds the upper threshold	storm-constrain control { block shutdown }	Optional By default, the storm constrain function is disabled.
Specify to send trap messages when the traffic detected exceeds the upper threshold or drops down below the lower threshold from a point higher than the upper threshold	storm-constrain enable trap	Optional By default, the system sends trap messages when the traffic detected exceeds the upper threshold or drops down below the lower threshold from a point higher than the upper threshold.

To do...	Use the command...	Remarks
Specify to send log when the traffic detected exceeds the upper threshold or drops down below the lower threshold from a point higher than the upper threshold	storm-constrain enable log	Optional By default, the system sends log when the traffic detected exceeds the upper threshold or drops down below the lower threshold from a point higher than the upper threshold.



- For network stability consideration, configure the interval for generating traffic statistics to a value that is not shorter than the default.
- The storm constrain function is applicable to multicast packets and broadcast packets on a port, and you can specify the upper and lower threshold for each of the two types of packets.

Maintaining and Displaying an Ethernet Port

To do...	Use the command...	Remarks
Display the current state of a specified port and related information	display interface [<i>interface-type</i> [<i>interface-number</i>]]	Available in any view
Display a summary of a specified port	display brief interface [<i>interface-type</i> [<i>interface-number</i>]] [{ begin include exclude } <i>text</i>]	Available in any view
Clear the statistics on a specified port	reset counters interface [<i>interface-type</i> [<i>interface-number</i>]]	Available in user view
Display the current ports of a specified type	display port { hybrid trunk }	Available in any view
Display the information about a manual port group or all the port groups	display port-group manual [all name <i>port-group-name</i>]	Available in any view
Display the information about the loopback function	display loopback-detection	Available in any view

17

PORT ISOLATION CONFIGURATION

When configuring port isolation, go to these sections for information you are interested in:

- "Introduction to Port Isolation" on page 157
- "Configuring an Isolation Group" on page 157
- "Displaying Isolation Groups" on page 158
- "Port Isolation Configuration Example" on page 158

Introduction to Port Isolation

To implement Layer 2 isolation, you can add different ports to different VLANs. However, this will waste the limited VLAN resource. With port isolation, the ports can be isolated within the same VLAN. Thus, you need only to add the ports to the isolation group to implement Layer 2 and Layer 3 isolation. This provides you with more secure and flexible networking schemes.

On the current device:

- A device supports only one isolation group that is created automatically by the system as Isolation Group 1. The user can neither delete the isolation group nor create other isolation groups.
- There is no restriction on the number of ports to be added to an isolation group.
- A port inside an isolation group and a port outside the isolation group can communicate with each other at Layer 2 and Layer 3. Ports of the isolation group cannot communicate with each other.

Configuring an Isolation Group

Adding a Port to an Isolation Group

Follow these steps to add a port to an isolation group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view or port group view	interface <i>interface-type</i> <i>interface-number</i> port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Use either command. Configured in Ethernet port view, the setting is effective on the current port only; configured in port group view, the setting is effective on all ports in the port group.

To do...	Use the command...	Remarks
Add a port to an isolation group as an ordinary port	port-isolate enable group <i>group-number</i>	Required No ports are added to the isolation group by default.

Displaying Isolation Groups

To do...	Use the command...	Remarks
Display an isolation group and its information	display port-isolate group	Available in any view

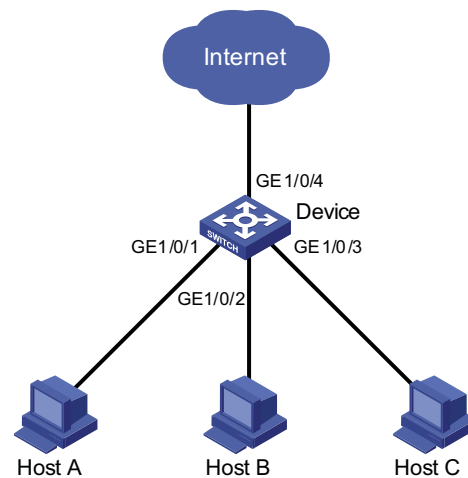
Port Isolation Configuration Example

Networking Requirement

- Users Host A, Host B, and Host C are connected to GigabitEthernet1/0/1, GigabitEthernet1/0/2, and GigabitEthernet1/0/3 of Device.
- Device is connected to an external network through Ethernet 2/0/4.
- GigabitEthernet1/0/1, GigabitEthernet1/0/2, GigabitEthernet1/0/3, and Ethernet 2/0/4 belong to the same VLAN. It is desired that Host A, Host B, and Host C cannot communicate with each other at Layer 2/Layer 3, but can access the external network.

Networking diagram

Figure 44 Network diagram for port isolation configuration



Configuration procedure

Add ports GigabitEthernet1/0/1, GigabitEthernet1/0/2 and GigabitEthernet1/0/3 to the isolation group.

```

<Device> system-view
[Device] interface GigabitEthernet1/0/1
[Device-GigabitEthernet1/0/1] port-isolate enable
[Device-GigabitEthernet1/0/1] quit
[Device] interface GigabitEthernet1/0/2
[Device-GigabitEthernet1/0/2] port-isolate enable
[Device-GigabitEthernet1/0/2] quit
  
```

```
[Device] interface GigabitEthernet1/0/3  
[Device-GigabitEthernet1/0/3] port-isolate enable
```

Display the information about the isolation group.

```
<Device> display port-isolate group  
Port-isolate group information:  
Uplink port support: No  
Group ID: 1  
GigabitEthernet1/0/1    GigabitEthernet1/0/2    GigabitEthernet1/0/3
```


18

LINK AGGREGATION OVERVIEW

This chapter covers these topics:

- “Link Aggregation” on page 161
- “Approaches to Link Aggregation” on page 162
- “Load Sharing in a Link Aggregation Group” on page 165
- “Service Loop Group” on page 165
- “Aggregation Port Group” on page 166

Link Aggregation

Link aggregation allows you to increase bandwidth by distributing traffic on the member ports in an aggregation group. In addition, it provides reliable connectivity because these member ports can dynamically back up each other.

LACP

Link Aggregation Control Protocol (LACP) is defined in IEEE 802.3ad. Link aggregation control protocol data unit (LACPDU) is used for exchanging information among LACP-enabled devices.

After LACP is enabled on a port, the port sends LACPDUs to notify the remote system of its system LACP priority, system MAC address, port LACP priority, port number, and operational key. Upon receipt of an LACPDU, the remote system compares the received information with the information received on other ports to determine the ports that can operate as selected ports. This allows the two systems to reach agreement on the states of the related ports

When aggregating ports, link aggregation control automatically assigns each port an operational key based on its rate, duplex mode, and other basic configurations. In an LACP aggregation group, all ports share the same operational key; in a manual or static LACP aggregation, the selected ports share the same operational key.

Consistency Considerations for Ports in an Aggregation

To participate in traffic sharing, member ports in an aggregation group must use the same configurations with respect to STP, QoS, GVRP, Q-in-Q, BPDU tunnel, VLAN, port attributes, MAC address learning, and so on as shown in the following table.

Table 29 Consistency considerations for ports in an aggregation

Category	Considerations
STP	<ul style="list-style-type: none"> State of port-level STP (enabled or disabled) Attribute of the link (point-to-point or otherwise) connected to the port Port path cost STP priority Maximum transmission rate Loop protection Root protection Port type (whether the port is an edge port)
QoS	<ul style="list-style-type: none"> Traffic policing Port rate limiting Strict priority (SP) queuing Weighted round robin (WRR) queuing Port priority Policy setting on the port Port trust mode
GVRP	<ul style="list-style-type: none"> GVRP state on ports (enabled or disabled) GVRP registration type GARP timers
Q-in-Q	<ul style="list-style-type: none"> State of Q-in-Q (enabled or disabled) Added outer VLAN tag Policy of appending outer VLAN tag according to inner VLAN IDs
BPDU tunnel	<ul style="list-style-type: none"> BPDU tunnel state on ports (enabled or disabled) BPDU tunnel state for STP on ports (enabled or disabled)
VLAN	<ul style="list-style-type: none"> VLANs carried on the port Default VLAN ID on the port Link type of the port, which can be trunk, hybrid, or access sub-net VLAN configuration protocol VLAN configuration VLAN tag configuration
Port attribute	<ul style="list-style-type: none"> Port rate Duplex mode Up/down state of the link Isolation group membership of the port
MAC address learning	<ul style="list-style-type: none"> Maximum number of MAC addresses that can be learned on the port

Approaches to Link Aggregation

Two ways are available for implementing link aggregation, as described in “Manual Link Aggregation” on page 163 and “Static LACP link aggregation” on page 164.

Manual Link Aggregation

Overview

Manual aggregations are created manually. Member ports in a manual aggregation are LACP-disabled.

Port states in a manual aggregation

In a manual aggregation group, ports are either selected or unselected. Selected ports can receive and transmit data frames whereas unselected ones cannot.

When setting the state of ports in a manual aggregation group, the system considers the following:

- The system selects the port with the highest priority in the up state as the reference port of the aggregation group. Port priority descends in the following order: full duplex/high speed, full duplex/low speed, half duplex/high speed, and half duplex/low speed. If multiple ports are of the same priority, the one with the lowest port number is the reference port.
- Ports in the up state with the same speed, duplex mode, link state, and basic configuration as the reference port become the candidates for selected ports, while the other ports become unselected ports.
- There is a limit on the number of selected ports in a manual aggregation group. If the number of selected-port candidates does not reach the limit, all the candidates become selected ports; if the number of candidates exceeds the limit, the candidates with lower port numbers become selected ports, while the other candidates become unselected ports.
- The selected port with the lowest port number serves as the master port of the aggregation group, and the other ports serve as the member ports of the aggregation group.
- If all the ports of an aggregation group are down, the port with the lowest port number is the master port. In this case, all of them are unselected ports.

In addition, unless the master port should be selected, a port that joins the group after the limit is reached will not be placed in selected state even if it should be in normal cases. This is to prevent the ongoing service on selected ports from being interrupted. You need to avoid the situation however as the selected/unselected state of a port may become different after a reboot.

Port Configuration Considerations in manual aggregation

As mentioned above, in a manual aggregation group, only ports with configurations consistent with those of the reference port can become selected. These configurations include port rate, duplex mode, link state, and other basic configurations, as described in "Consistency Considerations for Ports in an Aggregation" on page 161.

You need to maintain the basic configurations of these ports manually to ensure consistency. As one configuration change may involve multiple ports, this can become troublesome if you need to do that port by port. As a solution, you may add the ports into an aggregation port group where you can make configuration for all member ports.

When the configuration of some port in a manual aggregation group changes, the system does not remove the aggregation; instead, it re-sets the selected/unselected state of the member ports and re-selects a master port.

Static LACP link aggregation

Overview

Static aggregations are created manually. After you add a port to a static aggregation, LACP is enabled on it automatically.

Port states in static aggregation

In a static aggregation group, ports can be selected or unselected, where both can receive and transmit LACPDU's but only selected ports can receive and transmit data frames.

When setting the state of the ports in the local and remote static aggregation groups, the local and remote systems do the following:

- 1 Compare their system IDs to identify the higher priority system. (The system ID comprises LACP priority and system MAC address.)
- 2 First compare the system LACP priorities. The system with lower system LACP priority wins out.
- 3 If the system LACP priorities are the same, compare the system MAC addresses. The system with the smaller MAC address wins out.
- 4 Compare the port IDs on the higher priority system. (The port ID comprises port LACP priority and port number.)
- 5 Compare the port LACP priorities. The port with lower port LACP priority wins out.
- 6 If two ports with the same port LACP priority are present, compare their port numbers. The one with the smaller port ID wins out to become the reference port.
- 7 Select the candidates for selected ports. To be a candidate, a port must be in the up state with the same speed, duplex mode, link state, and basic configuration as the reference port; in addition, their peer ports on the other system must have the same configuration. All the ports but the selected-port candidates become unselected.
- 8 As there is a limit on the number of selected ports, not all selected-port candidates can become selected ports. Before the limit is reached, all the candidates are set to the selected state. When the limit is reached, the candidates with lower port numbers are set to the selected state while the other candidates are set to the unselected state. At the same time, the other system gets aware of the state change of the ports on the higher priority system and thus sets the state of the corresponding local ports.
- 9 Set the selected port with the lowest port number as the master port in the aggregation group on each system.

Port configuration considerations in static aggregation

Like in a manual aggregation group, in a static LACP aggregation group, only ports with configurations consistent with those of the reference port can become selected. These configurations include port rate, duplex mode, link state and other basic configurations described in "Consistency Considerations for Ports in an Aggregation" on page 161.

You need to maintain the basic configurations of these ports manually to ensure consistency. As one configuration change may involve multiple ports, this can become troublesome if you need to do that port by port. As a solution, you may add the ports into an aggregation port group where you can make configuration for all member ports.

When the configuration of some port in a static aggregation group changes, the system does not remove the aggregation; instead, it re-sets the selected/unselected state of the member ports and re-selects a master port.

Load Sharing in a Link Aggregation Group

Link aggregation groups fall into load sharing aggregation groups and non-load sharing aggregation groups depending on their support to load sharing.

A load sharing aggregation group can contain at least one selected port but a non-load sharing aggregation group can contain only one.

Link aggregation groups perform load sharing depending on availability of hardware resources. When hardware resources are available, link aggregation groups created containing at least two selected ports perform load sharing, while link aggregation groups created with only one selected port does not perform load sharing. After hardware resources become depleted, link aggregation groups created work in non-load sharing mode.

Load sharing is implemented through the selected ports in an aggregation group. However, the way of selecting forwarding ports varies by packet type:

- For a Layer-2 unicast packet with a known destination MAC address, if the packet carries an IP datagram, the switch selects the forwarding port according to the source IP address and destination IP address; otherwise, the switch selects the forwarding port according to the source MAC address and destination MAC address.
- For a unicast IP packet with a known destination IP address, the switch selects the forwarding port according to the source IP address and the destination IP address of the packet.
- For a Layer-2 multicast packet with a known destination MAC address, the switch selects the forwarding port according to the source MAC address, the destination MAC address, and the receiving port of the packet.
- For a Layer-3 multicast packet with a known IP address, the switch selects the forwarding port according to the source IP address, the destination IP address, and the receiving port of the packet.
- For an unknown unicast/multicast/broadcast packet, the switch selects the forwarding port according to the source MAC address, the destination MAC address, and the receiving port of the packet.



When only one selected port remains in a load sharing aggregation group, the group keeps working in the load sharing mode.

Service Loop Group

You can create a service loop group by creating a manual aggregation group of service-loop ports first and then specifying which services can be redirected for the

group. At present, you may specify to redirect four types of services, IPv6 (IPv6 unicast), IPv6mc (IPv6 multicast), tunnel, and MPLS.



Currently, the the Switch 4800G support to redirect tunnel services only.

After creating a service-loop group, assign ports that support its service type to the group considering the following:

- These ports can be configured only with the physical configuration such as speed and duplex mode, QoS, and ACL. Other conflicting configurations, such as STP cannot be configured.
- These ports must belong to VLAN 1.

After assigning a port to a service-loop group, you may configure it with other non-conflicting settings, such as QoS.

If this group is performing load sharing, it continues to function in this way even after all selected ports but one are removed to ensure ongoing service.

Aggregation Port Group

As mentioned earlier, in a manual or static aggregation group, a port can be selected only when its configuration is the same as that of the reference port in terms of duplex/speed pair, link state, and other basic configurations. Their configuration consistency requires administrative maintenance, which is troublesome after you change some configuration.

To simplify configuration, port-groups are provided allowing you to configure for all ports in individual groups at one time. One example of port-groups is aggregation port group.

Upon creation or removal of a link aggregation group, an aggregation port-group which cannot be administratively created or removed is automatically created or removed. In addition, you can only assign/remove a member port to/from an aggregation port-group by assigning/removing it from the corresponding link aggregation group.

For more information about port-groups, refer to “Configuring a Port Group” on page 150.

19

LINK AGGREGATION CONFIGURATION

When configuring link aggregation, go to these sections for information you are interested in:

- “Configuring Link Aggregation” on page 167
- “Displaying and Maintaining Link Aggregation” on page 169
- “Link Aggregation Configuration Example” on page 170

Configuring Link Aggregation

This section covers these topics:

- “Configuring a Manual Link Aggregation Group” on page 167
- “Configuring a Static LACP Link Aggregation Group” on page 168
- “Configuring an Aggregation Group Name” on page 168
- “Configuring a Service Loop Group” on page 169
- “Entering Aggregation Port Group View” on page 169

Configuring a Manual Link Aggregation Group

Follow these steps to create a manual aggregation group and add an Ethernet port to it:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Create a manual aggregation group	link-aggregation group <i>agg-id</i> mode manual	Required
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	--
Assign the Ethernet port to the aggregation group	port link-aggregation group <i>agg-id</i>	Required



- *You can create a manual aggregation group by changing the type of an existing static aggregation group. When you create a manual aggregation group in this way and the static aggregation group contains ports, LACP is disabled on the ports after the manual aggregation group is created.*
- *An aggregation group cannot contain the following ports: RRPP-enabled ports, ports configured with static MAC addresses or black hole MAC addresses, voice VLAN-enabled ports, or 802.1x-enabled ports.*
- *After you remove a manual aggregation group, all the ports in the group are dismissed from it.*
- *For a manual aggregation group containing only one port, the only way to remove the port from it is to remove the aggregation group.*

- To make an aggregation group to function properly, make sure the selected states of the ports on the both sides of the same link are the same.

Configuring a Static LACP Link Aggregation Group

Follow these steps to configure a static aggregation group:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Configure the system LACP priority	lACP system-priority <i>system-priority</i>	Optional 32768 by default. Changing system LACP priority can affect the selected/unselected state of the ports in the group.
Create a static LACP aggregation group	link-aggregation group <i>agg-id mode static</i>	Required
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	--
Configure the port LACP priority	lACP port-priority <i>port-priority</i>	Optional 32768 by default. Changing port LACP priority can affect the selected/unselected state of the ports in the group.
Assign the Ethernet port to the aggregation group	port link-aggregation group <i>agg-id</i>	Required



- You can create a static aggregation group by changing the type of an existing manual link aggregation group that contains no port.
- An aggregation group cannot contain the following ports: RRPP-enabled ports, ports configured with static MAC addresses or black hole MAC addresses, voice VLAN-enabled ports, or 802.1x-enabled ports.
- After you remove a static aggregation group, all the ports in the group are dismissed from it, and LACP is disabled on the ports.
- For a static LACP aggregation group containing only one port, the only way to remove the port from the aggregation group is to remove the aggregation group.



When making configuration, be aware that after a load-balancing aggregation group changes to a non-load balancing group due to resources exhaustion, either of the following may happen

- Forwarding anomaly resulted from inconsistency of the two ends in the number of selected ports.
- Some protocols such as GVRP malfunction because the state of the remote port connected to the master port is unselected.

Configuring an Aggregation Group Name

Follow these steps to configure a name for an aggregation group:

To do...	Use the command...	Remarks
Enter system view	system-view	--

To do...	Use the command...	Remarks
Configure a name for a link aggregation group	link-aggregation group <i>agg-id</i> description <i>agg-name</i>	Required None is configured by default.

Configuring a Service Loop Group

Follow these steps to configure a service loop group:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Create a manual aggregation group	link-aggregation group <i>agg-id</i> mode manual	Required
Specify the aggregation group as a service loop group that is of specific type	link-aggregation group <i>agg-id</i> service-type tunnel	Required
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	--
Add the Ethernet port to the aggregation group	port link-aggregation group <i>agg-id</i>	Required



- You can remove any service loop group except those that are currently referenced by modules.
- For a service loop group containing only one port, the only way to remove the port from it is to remove the service loop group.

Entering Aggregation Port Group View

In aggregation port group view, you can make configuration for all the member ports in a link aggregation group at one time.

Follow these steps to enter aggregation port group view:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter aggregation port group view	port-group aggregation <i>agg-id</i>	--



CAUTION: In aggregation port group view, you can configure aggregation related settings such as STP, VLAN, QoS, GVRP, Q-in-Q, BPDU tunnel, MAC address learning, but cannot add or remove member ports.

Displaying and Maintaining Link Aggregation

To do...	Use the command...	Remarks
Display the local system ID	display lacp system-id	Available in any view
Display detailed information about link aggregation for the specified port or ports	display link-aggregation interface <i>interface-type</i> <i>interface-number</i> [to <i>interface-type</i> <i>interface-number</i>]	Available in any view
Display information about the specified or all service loop groups	display link-aggregation service-type [<i>agg-id</i>]	Available in any view
Display summaries for all link aggregation groups	display link-aggregation summary	Available in any view

To do...	Use the command...	Remarks
Display detailed information about specified or all link aggregation groups	display link-aggregation verbose [<i>agg-id</i>]	Available in any view
Clear the statistics about LACP for specified or all ports	reset lacp statistics [interface <i>interface-type interface-number</i> [to <i>interface-type interface-number</i>]]	Available in user view

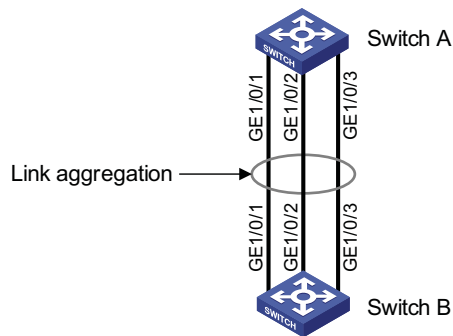
Link Aggregation Configuration Example

Network requirements

- Switch A aggregates ports GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to form one link connected to Switch B and performs load sharing among these ports.
- Create a tunnel service-loop group and add port GigabitEthernet 1/0/1 to the group.

Network diagram

Figure 45 Network diagram for link aggregation configuration



Configuration procedure



This example only describes how to configure link aggregation on Switch A. To achieve link aggregation, do the same on Switch B.

1 In manual aggregation approach

Create manual aggregation group 1.

```
<SwitchA> system-view
[SwitchA] link-aggregation group 1 mode manual
```

Add ports GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to the group.

```
[SwitchA] interface GigabitEthernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] port link-aggregation group 1
[SwitchA-GigabitEthernet1/0/1] interface GigabitEthernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] port link-aggregation group 1
[SwitchA-GigabitEthernet1/0/2] interface GigabitEthernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] port link-aggregation group 1
```

2 In static aggregation approach

Create static aggregation group 1.

```
<SwitchA> system-view  
[SwitchA] link-aggregation group 1 mode static
```

Add ports GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to the group.

```
[SwitchA] interface GigabitEthernet 1/0/1  
[SwitchA-GigabitEthernet1/0/1] port link-aggregation group 1  
[SwitchA-GigabitEthernet1/0/1] interface GigabitEthernet 1/0/2  
[SwitchA-GigabitEthernet1/0/2] port link-aggregation group 1  
[SwitchA-GigabitEthernet1/0/2] interface GigabitEthernet 1/0/3  
[SwitchA-GigabitEthernet1/0/3] port link-aggregation group 1
```

3 Configure a service loop group

Create a manual aggregation group.

```
<SwitchA> system-view  
[SwitchA] link-aggregation group 1 mode manual
```

Specify this group to be a tunnel service loop group.

```
[SwitchA] link-aggregation group 1 service-type tunnel
```

Assign port GigabitEthernet 1/0/1 to the service loop group.

```
[SwitchA] interface GigabitEthernet 1/0/1  
[SwitchA-GigabitEthernet1/0/1] undo stp  
[SwitchA-GigabitEthernet1/0/1] port link-aggregation group 1
```


20

MAC ADDRESS TABLE MANAGEMENT CONFIGURATION

When configuring MAC address table management, go to these sections for information you are interested in:

- “Introduction to MAC Address Table” on page 173
- “Configuring MAC Address Table Management” on page 174
- “Displaying and Maintaining MAC Address Table Management” on page 176
- “MAC Address Table Management Configuration Example” on page 176



This manual covers only static, dynamic and blackhole MAC address table management. For the management of multicast MAC address table management, refer to “Multicast Routing and Forwarding Configuration” on page 701.

Introduction to MAC Address Table

A switch maintains a MAC address table for frame forwarding. Each entry in this table contains the MAC address of a connected device, to which port this device is connected and to which VLAN the port belongs.

A MAC address table consists of two types of entries: static and dynamic. Static entries are manually configured and never age out. Dynamic entries can be manually configured or dynamically learned and may age out.

The following is how a switch learns a MAC address after it receives a frame from a port, port A for example:

- 1 Check the frame for the source MAC address (MAC-SOURCE for example), and consider frames with destination MAC address MAC-SOURCE be forwarded through port A.
- 2 Look up the MAC address table for an entry corresponding to the MAC address and do the following:
- 3 If an entry is found for the MAC address, update the entry.
- 4 If no entry is found, add an entry containing the MAC address and port A.

When receiving a frame destined for MAC A, the switch then looks up the MAC address table and forwards it from port A.

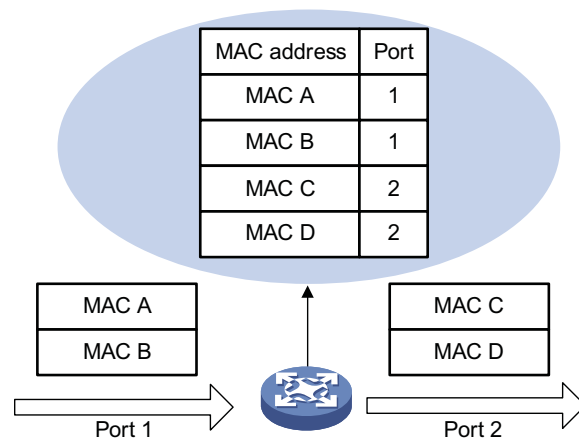


Dynamically learned MAC addresses cannot overwrite static MAC address entries, but the latter can overwrite the former.

As shown in Figure 46, when forwarding a frame, the switch looks up the MAC address table. If an entry is available for the destination MAC address, the switch forwards the frame directly from the hardware. If not, it does the following:

- 1 Broadcast the frame.
- 2 After the frame reaches the destination, the destination sends back a response with its MAC address. (If no response is received, the frame will be dropped.)
- 3 Upon receipt of the response, the device adds an entry in the MAC address table, indicating from which port the frames destined for the MAC address should be sent.
- 4 Forward subsequent frames destined for the same MAC address directly from the hardware.
- 5 Discard the frames which cannot reach the destination MAC address.

Figure 46 Forward frames using the MAC address table



Configuring MAC Address Table Management

This section covers these topics:

- “Configuring MAC Address Entries” on page 174
- “Configuring MAC Address Aging Timer” on page 175
- “Configuring the Maximum Number of MAC Addresses an Ethernet Port or a Port Group Can Learn” on page 175

Configuring MAC Address Entries

Follow these steps to add, modify, or remove entries in the MAC address table:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Add/modify a MAC address entry	mac-address blackhole <i>mac-address</i> vlan <i>vlan-id</i> mac-address { dynamic static } <i>mac-address</i> interface <i>interface-type</i> <i>interface-number</i> vlan <i>vlan-id</i>	Required
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	-
Add/modify MAC address entries in the specified port view	mac-address { dynamic static } <i>mac-address</i> vlan <i>vlan-id</i>	Required



Do not configure a static or dynamic MAC address entry on an aggregation port.

Configuring MAC Address Aging Timer

The MAC address table on your device is available with an aging mechanism for dynamic entries to prevent its resources from being exhausted. Set the aging timer appropriately: a long aging interval may cause the MAC address table to retain outdated entries and fail to accommodate latest network changes; a short interval may result in removal of valid entries and hence unnecessary broadcasts which may affect device performance.

Follow these steps to configure the MAC address aging timer:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the aging timer for dynamic MAC address entries	mac-address timer { aging seconds no-aging }	Optional 300 seconds by default.



The MAC address aging timer takes effect globally on dynamic MAC address entries (learned or administratively configured) only.

Configuring the Maximum Number of MAC Addresses an Ethernet Port or a Port Group Can Learn

To prevent a MAC address table from getting so large that it may degrade forwarding performance, you may restrict the number of MAC addresses that can be learned. One approach is to do this on a per-port or port group basis.

Follow these steps to configure the maximum number of MAC addresses that an Ethernet port or port group can learn:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port or port group view	interface <i>interface-type interface-number</i> port-group { aggregation agg-id manual port-group-name }	Required Use either command to configure on a port or ports in a group.
Configure the maximum number of MAC addresses that can be learned on an Ethernet port or port group.	mac-address max-mac-count <i>count</i>	Required The maximum number of MAC addresses that can be learned on a port or port group is not configured by default.

Displaying and Maintaining MAC Address Table Management

To do...	Use the command...	Remarks
Display MAC address table information	display mac-address blackhole [<i>vlan vlan-id</i>] [<i>count</i>] display mac-address [<i>mac-address</i> [<i>vlan vlan-id</i>]] [dynamic static] [interface <i>interface-type interface-number</i>] [<i>vlan vlan-id</i>] [<i>count</i>]	Available in any view
Display the aging timer for dynamic MAC address entries	display mac-address aging-time	

MAC Address Table Management Configuration Example

Network requirements

Log onto your device from the Console port to configure MAC address table management as follows:

- Set the aging timer to 500 seconds for dynamic MAC address entries.
- Add a static entry 000f-e235-dc71 for port GigabitEthernet 1/0/1 in VLAN 1.

Configuration procedure

Add a static MAC address entry.

```
<Sysname> system-view
[Sysname] mac-address static 000f-e235-dc71 interface GigabitEthernet 1/0/1 vlan 1
```

Set the aging timer for dynamic MAC address entries to 500 seconds.

```
[Sysname] mac-address timer aging 500
```

Display the MAC address entry for port GigabitEthernet 1/0/1.

```
[Sysname] display mac-address interface GigabitEthernet 1/0/1
MAC ADDR          VLAN ID  STATE          PORT INDEX          AGING TIME(s)
000f-e235-dc71    1        Config static   GigabitEthernet 1/0/1 NOAGED

--- 1 mac address(es) found ---
```

21

IP SOURCE GUARD CONFIGURATION

When configuring IP Source Guard, go to these sections for information you are interested in:

- "IP Source Guard Overview" on page 177
- "Configuring a Static Binding Entry" on page 177
- "Configuring Dynamic Binding Function" on page 178
- "Displaying IP Source Guard" on page 178
- "IP Source Guard Configuration Examples" on page 178
- "Troubleshooting" on page 182

IP Source Guard Overview

By filtering packets on a per-port basis, IP source guard prevents packets with illegal IP addresses and MAC addresses from traveling through, improving the network security. After receiving a packet, the port looks up the key attributes (including IP address, MAC address and VLAN tag) of the packet in the binding entries of the IP source guard. If there is a matching entry, the port will forward the packet. Otherwise, the port will abandon the packet.

IP source guard filters packets based on the following types of binding entries:

- IP-port binding entry
- MAC-port binding entry
- IP-MAC-port binding entry

You can manually set static binding entries, or use DHCP Snooping to provide dynamic binding entries. Binding is on a per-port basis. After a binding entry is configured on a port, it is effective only to the port, instead of other ports.



CAUTION: *IP source guard and aggregation group configuration are mutually exclusive.*

Configuring a Static Binding Entry

Follow these steps to configure a static binding entry:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-

To do...	Use the command...	Remarks
Configure a static binding entry	user-bind { ip-address <i>ip-address</i> ip-address <i>ip-address</i> mac-address <i>mac-address</i> mac-address <i>mac-address</i> }	Required No static binding entry exists by default.



- The system does not support repeatedly configuring a binding entry to one port. A binding entry can be configured to multiple ports.
- In a valid binding entry, the MAC address cannot be all 0s, all Fs (a broadcast address), or a multicast address, and the IP address can only be a Class A, Class B, or Class C address and can be neither 127.x.x.x nor 0.0.0.0.

Configuring Dynamic Binding Function

After the dynamic binding function is enabled on a port, IP source guard will receive and process corresponding DHCP Snooping entries, which contain such information as MAC address, IP address, VLAN tag, port information or entry type. It adds the obtained information to the dynamic binding entries to enable the port to filter packets according to the binding entries.

Follow these steps to configure dynamic binding function:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure dynamic binding function	ip check source { ip-address ip-address mac-address mac-address }	Required Not configured by default

Displaying IP Source Guard

To do...	Use the command...	Remarks
Display information about static binding entries	display user-bind [interface <i>interface-type</i> <i>interface-number</i> ip-address <i>ip-address</i> mac-address <i>mac-address</i>]	Available in any view
Display information about dynamic binding entries	display ip check source [interface <i>interface-type</i> <i>interface-number</i> ip-address <i>ip-address</i> mac-address <i>mac-address</i>]	Available in any view

IP Source Guard Configuration Examples

Static Binding Entry Configuration Example

Network requirements

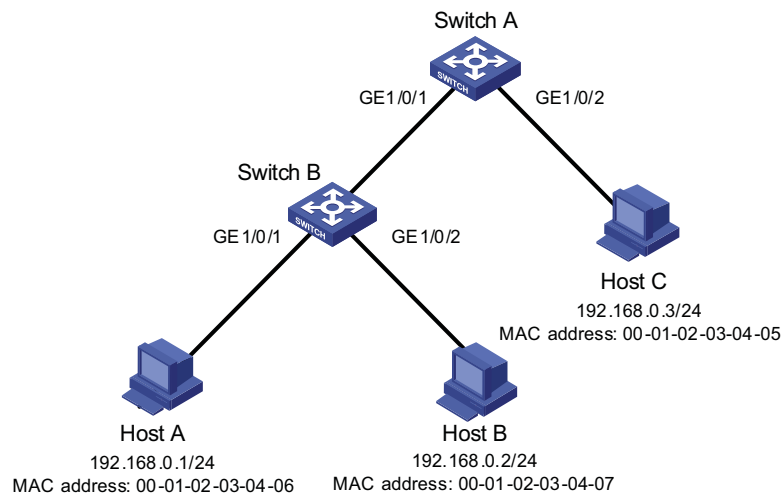
As shown in Figure 47, switches A and B and Hosts A, B and C are on an Ethernet. Host A and Host B are connected to ports GigabitEthernet1/0/1 and GigabitEthernet1/0/2 of Switch B respectively, Host C is connected to port GigabitEthernet1/0/2 of Switch A, while Switch B is connected to port GigabitEthernet1/0/1 of Switch A.

Detailed requirements are as follows:

- On port GigabitEthernet1/0/2 of Switch A, only IP packets with the source MAC address of 00-01-02-03-04-05 and the source IP address of 192.168.0.3 can pass.
- On port GigabitEthernet1/0/1 of Switch A, only IP packets with the source MAC address of 00-01-02-03-04-06 and the source IP address of 192.168.0.1 can pass.
- On port GigabitEthernet1/0/1 of Switch B, only IP packets with the source MAC address of 00-01-02-03-04-06 and the source IP address of 192.168.0.1 can pass.
- On port GigabitEthernet1/0/2 of Switch B, only IP packets with the source MAC address of 00-01-02-03-04-07 and the source IP address of 192.168.0.2 can pass.

Network diagram

Figure 47 Network diagram for configuring static binding entries



Configuration procedure

1 Configure Switch A

Configure the IP addresses of various interfaces (omitted).

Configure port GigabitEthernet1/0/2 of Switch A to allow only IP packets with the source MAC address of 00-01-02-03-04-05 and the source IP address of 192.168.0.3 to pass.

```

<SwitchA> system-view
[SwitchA] interface GigabitEthernet 1/0/2
[SwitchA-GigabitEthernet1/0/2] user-bind ip-address 192.168.0.3 mac-address 0001-0203-0405
[SwitchA-GigabitEthernet1/0/2] quit
  
```

Configure port GigabitEthernet1/0/1 of Switch A to allow only IP packets with the source MAC address of 00-01-02-03-04-06 and the source IP address of 192.168.0.1 to pass.

```

[SwitchA] interface GigabitEthernet1/0/1
[SwitchA-GigabitEthernet1/0/1] user-bind ip-address 192.168.0.1 mac-address 0001-0203-0406
  
```

2 Configure Switch B

Configure the IP addresses of various interfaces (omitted).

Configure port GigabitEthernet1/0/1 of Switch B to allow only IP packets with the source MAC address of 00-01-02-03-04-06 and the source IP address of 192.168.0.1 to pass.

```
<SwitchB> system-view
[SwitchB] interface GigabitEthernet1/0/1
[SwitchB-GigabitEthernet1/0/1] user-bind ip-address 192.168.0.1 mac-address 0001-0203-0406
[SwitchB-GigabitEthernet1/0/1] quit
```

Configure port GigabitEthernet1/0/2 of Switch B to allow only IP packets with the source MAC address of 00-01-02-03-04-07 and the source IP address of 192.168.0.2 to pass.

```
[SwitchB] interface GigabitEthernet1/0/2
[SwitchB-GigabitEthernet1/0/2] user-bind ip-address 192.168.0.2 mac-address 0001-0203-0407
```

3 Verify the configuration

On Switch A, static binding entries are configured successfully.

```
<SwitchA> display user-bind
The following user address bindings have been configured:
MAC                IP                Vlan  Port                Status
0001-0203-0405    192.168.0.3      N/A   GigabitEthernet1/0/2  Static
0001-0203-0406    192.168.0.1      N/A   GigabitEthernet1/0/1  Static
-----2 binding entries queried, 2 listed-----
```

On Switch B, static binding entries are configured successfully.

```
<SwitchB> display user-bind
The following user address bindings have been configured:
MAC                IP                Vlan  Port                Status
0001-0203-0406    192.168.0.1      N/A   GigabitEthernet1/0/1  Static
0001-0203-0407    192.168.0.2      N/A   GigabitEthernet1/0/2  Static
-----2 binding entries queried, 2 listed-----
```

Dynamic Binding Function Configuration Example

Network requirements

Switch A connects to Client A and the DHCP Server through GigabitEthernet1/0/1 and GigabitEthernet1/0/2 respectively. DHCP Snooping is enabled on Switch A.

Detailed requirements are as follows:

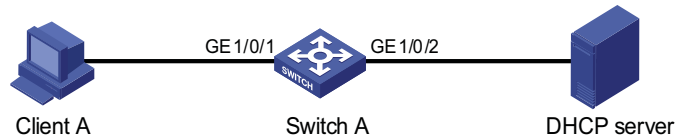
- Client A with the MAC address of 00-01-02-03-04-06 obtains an IP address through the DHCP Server.
- On Switch A, create the DHCP Snooping entry of Client A.
- On port GigabitEthernet1/0/1 of Switch A, enable dynamic binding function to prevent attacks from using forged IP addresses to attack the server.



For detailed configuration of DHCP Server, refer to "DHCP Server Configuration" on page 797.

Network diagram

Figure 48 Network diagram for configuring dynamic binding



Configuration procedure

1 Configure Switch A

Configure dynamic binding on port GigabitEthernet1/0/1.

```

<SwitchA> system-view
[SwitchA] interface GigabitEthernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] ip check source ip-address mac-address
[SwitchA-GigabitEthernet1/0/1] quit
  
```

Enable DHCP snooping on Switch A.

```

[SwitchA] dhcp-snooping
  
```

Configure port GigabitEthernet1/0/2 connected to the DHCP server as a trusted port.

```

[SwitchA] interface GigabitEthernet1/0/2
[SwitchA-GigabitEthernet1/0/2] dhcp-snooping trust
[SwitchA-GigabitEthernet1/0/2] quit
  
```

2 Verify the configuration

Display the dynamic binding entries that port GigabitEthernet1/0/1 has obtained from DHCP Snooping.

```

<SwitchA> display ip check source
The following user address bindings have been configured:
MAC          IP          Vlan  Port          Status
0001-0203-0406  192.168.0.1  1    GigabitEthernet1/0/1  DHCP-SNP
-----1 binding entries queried, 1 listed-----
  
```

Display the dynamic entries of DHCP Snooping and check it is identical with the dynamic entries that port GigabitEthernet1/0/1 has obtained.

```

<SwitchA> display dhcp-snooping
DHCP Snooping is enabled.
The client binding table for all untrusted ports.
Type : D--Dynamic , S--Static
Type IP Address      MAC Address      Lease      VLAN  Interface
==== =====
D   192.168.0.1      0001-0203-0406  86335     1    GigabitEthernet1/0/1
  
```

As you see, port GigabitEthernet1/0/1 has obtained the dynamic entries generated by DHCP Snooping after it is configured with dynamic binding function.

Troubleshooting

Failed to Configure Static Binding Entries and Dynamic Binding Function

Symptom

Configuring static binding entries and dynamic binding function fails on a port.

Analysis

IP Source Guard is not supported on the port which has joined an aggregation group. Neither static binding entries nor dynamic binding function can be configured on the port which has joined an aggregation group.

Solution

Remove the port from the aggregation group.

22

DLDP CONFIGURATION

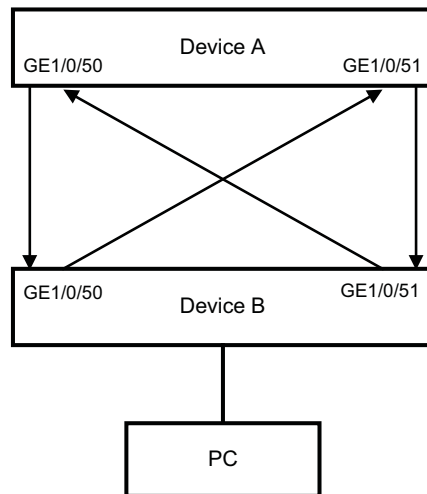
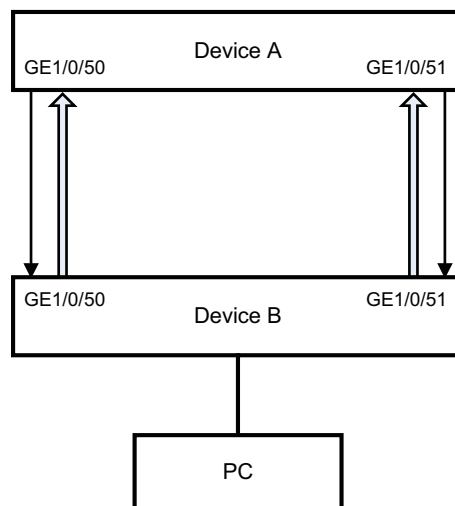
When performing DLDP configuration, go to these sections for information you are interested in:

- "Overview" on page 183
- "DLDP Configuration Task List" on page 190
- "Enabling DLDP" on page 190
- "Setting DLDP Mode" on page 191
- "Setting the Interval for Sending Advertisement Packets" on page 191
- "Setting the DelayDown Timer" on page 191
- "Setting the Port Shutdown Mode" on page 192
- "Configuring DLDP Authentication" on page 192
- "Resetting DLDP State" on page 193
- "Displaying and Maintaining DLDP" on page 193
- "DLDP Configuration Example" on page 193
- "Troubleshooting" on page 195

Overview

A special kind of links, namely, unidirectional links, may occur in a network. When a unidirectional link appears, the local device can receive packets from the peer device through the link layer, but the peer device cannot receive packets from the local device. Unidirectional link can cause problems such as loops in a Spanning Tree Protocol (STP) enabled network.

As for fiber links, two kinds of unidirectional links exist. One occurs when fibers are cross-connected, as shown in Figure 49. The other occurs when one end of a fiber is not connected or one fiber of a fiber pair gets disconnected, as illustrated by the hollow arrows in Figure 50.

Figure 49 Unidirectional fiber link: cross-connected fiber**Figure 50** Unidirectional fiber link: fiber not connected or disconnected**DLDP Introduction**

Device Link Detection Protocol (DLDP) can detect the link status of a fiber cable or twisted pair. On detecting a unidirectional link, DLDP can shut down the related port automatically or prompt users to take measures as configured to avoid network problems.

As a data link layer protocol, DLDP cooperates with physical layer protocols to monitor the link status of a device. The auto-negotiation mechanism provided by physical layer protocols detects physical signals and faults. DLDP, however, performs operations such as identifying peer devices, detecting unidirectional links, and shutting down unreachable ports. The cooperation of physical layer protocols and DLDP ensures that physical/logical unidirectional links be detected and shut down. For a link with the devices on the both sides of it operating properly, DLDP checks to see if the cable is connected correctly and if packets can be exchanged between the two devices. Note that DLDP is not implemented through auto-negotiation.

DLDP Fundamentals DLDP link states

A device is in one of these DLDP link states: Initial, Inactive, Active, Advertisement, Probe, Disable, and DelayDown, as described in Table 30.

Table 30 DLDP link states

State	Description
Initial	This state indicates that DLDP is not enabled.
Inactive	This state indicates that DLDP is enabled but the link is down.
Active	This state indicates that: <ul style="list-style-type: none"> ■ DLDP is enabled and the link is up. ■ The neighbor entries are cleared.
Advertisement	This state indicates that a device can communicate normally with all its neighbors in both directions or DLDP remains in active state for more than five seconds. It is the normal state where no unidirectional link is detected.
Probe	A device enters this state if it receives a packet from an unknown neighbor. In this state, DLDP sends packets to check whether the link is a unidirectional link. After a device enters this state, the probe sending timer is triggered, and an echo waiting timer is triggered for each neighbor to be detected.
Disable	A device enters this state when: <ul style="list-style-type: none"> ■ A unidirectional link is detected. ■ The contact with a neighbor in enhanced mode gets lost. <p>In this state, no DLDP packet is sent or accepted.</p>
DelayDown	A device in the Active, Advertisement, or Probe DLDP link state transits to this state rather than remove the corresponding neighbor entry and transits to the Inactive state when it detects a port-down event. When a device transits to this state, the DelayDown timer is triggered.

DLDP timers

Table 31 DLDP timers

DLDP timer	Description
Active timer	Determines the Interval to send Advertisement packets with RSY tag, which defaults to 1 second. When a device transits to the active DLDP link state, it sends Advertisement packets with RSY tag according to this timer. The maximum number of this type of packets that can be sent successively is 5.
Advertisement timer	Determines the interval to send advertisement packets, which defaults to 5 seconds.
Probe timer	Determines the interval to send Probe packets, which defaults to 0.5 seconds. The maximum number of this type of packets that can be sent successively is 10.
Echo timer	This timer is set to 10 seconds and is triggered when a device transits to the Probe state or an enhanced detect is launched. When the Echo waiting timer expires and no Echo packet is received from a neighbor device, the link is set as a unidirectional link and the device transits to the Disable state. In this case, the device sends Disable packets, prompts the user to shut down the port or shuts down the port automatically (depending on the DLDP down mode configured), and removes the corresponding neighbor entries.

Table 31 DLDP timers

DLDP timer	Description
Entry timer	<p>When a new neighbor joins, a neighbor entry is created and the corresponding entry timer is triggered. And when a DLDP packet is received, the device updates the corresponding neighbor entry and the entry aging timer.</p> <p>In the normal mode, if no packet is received from a neighbor when the corresponding entry aging timer expires, DLDP sends advertisement packets with RSY tags and removes the neighbor entry.</p> <p>In the enhanced mode, if no packet is received from a neighbor when the Entry timer expires, DLDP triggers the enhanced timer.</p> <p>The setting of an Entry timer is three times that of the Advertisement timer.</p>
Enhanced timer	<p>In the enhanced mode, this timer is triggered if no packet is received from a neighbor when the entry aging timer expires. Enhanced timer is set to 10 seconds.</p> <p>After the Enhanced timer is triggered, the device sends up to eight probe packets to the neighbor at a frequency of one packet per second. If no Echo packet is received from the neighbor when the Echo timer expires, the link is set as a unidirectional link and the device transits to the Disable state. In this case, the device sends Disable packets, prompts the user to shut down the port or shuts down the port automatically (depending on the DLDP down mode configured), and removes the corresponding neighbor entries.</p>
DelayDown timer	<p>A device in the Active, Advertisement, or Probe DLDP link state transits to DelayDown state rather than removes the corresponding neighbor entry and transits to the Inactive state when it detects a port-down event.</p> <p>When a device transits to this state, the DelayDown timer is triggered. The setting of the timer ranges from 1 to 5 (in seconds). A device in DelayDown state only responds to port-up events.</p> <p>A device in the DelayDown state resumes its original DLDP state if it detects a port-up event before the DelayDown timer expires. Otherwise, it removes the corresponding DLDP neighbor information and transits to the Inactive state.</p>
RecoverProbe timer	<p>Determines the interval to RecoverProbe packets, which are used to detect whether a unidirectional link is restored. This timer is set to 2 seconds.</p>

DLDP mode

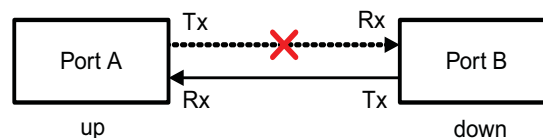
DLDP can operate in two modes: normal mode and enhanced mode, as described below.

- In normal DLDP mode, when an entry timer expires, the device removes the corresponding neighbor entry and sends an Advertisement packet with RSY tag.
- In enhanced DLDP mode, when an entry timer expires, the Enhanced timer is triggered and the device sends up to eight Probe packets at a frequency of one packet per second to test the neighbor. If no Echo packet is received from the neighbor when the Echo timer expires, the device transits to the Disable state.

Table 32 DLDAP mode and neighbor entry aging

DLDP mode	Detecting a neighbor after the corresponding neighbor entry ages out	Removing the neighbor entry immediately after the Entry timer expires	Triggering the Enhanced timer after an Entry timer expires
Normal DLDP mode	No	Yes	No
Enhanced DLDP mode	Yes	No	Yes

The enhanced DLDP mode is designed for addressing black holes. It prevents the cases where one end of a link is up and the other is down. If you configure the speed and the duplex mode by force on a device, the situation shown in Figure 51 may occur, where Port B is actually down but the state of Port B cannot be detected by common data link protocols, so Port A is still up. In enhanced DLDP mode, however, Port A tests Port B after the Entry timer concerning Port B expires. Port A then transits to the Disable state if it receives no Echo packet from Port A when the Echo timer expires. As Port B is physically down, it is in the Inactive DLDP state.

Figure 51 A case for Enhanced DLDP mode

- In normal DLDP mode, only fiber cross-connected unidirectional links (as shown in Figure 49) can be detected.
- In enhanced DLDP mode, two types of unidirectional links can be detected. One is fiber cross-connected links (as shown in Figure 49). The other refers to fiber pairs with one fiber not connected or disconnected (as shown in Figure 50). To detect unidirectional links that are of the latter type, you need to configure the ports to operate at specific speed and in full duplex mode. Otherwise, DLDP cannot take effect. When a fiber of a fiber pair is not connected or gets disconnected, the port that can receive optical signals is in Disable state; the other port is in Inactive state.

DLDP authentication mode

You can prevent network attacks and illegal detect through DLDP authentication. Three DLDP authentication modes exist, as described below.

- Non-authentication. In this mode, the sending side sets the Authentication field and the Authentication type field of DLDP packets to 0. The receiving side checks the values of the two fields of received DLDP packets and drops the packets with the two fields conflicting with the corresponding local configuration.
- Plain text authentication. In this mode, before sending a DLDP packet, the sending side sets the Authentication field to the password configured in plain text and sets the Authentication type field to 1. The receiving side checks the values of the two fields of received DLDP packets and drops the packets with the two fields conflicting with the corresponding local configuration.

- MD5 authentication. In this mode, before sending a packet, the sending side encrypts the user configured password using MD5 algorithm, assigns the digest to the Authentication field, and sets the Authentication type field to 2. The receiving side checks the values of the two fields of received DLDP packets and drops the packets with the two fields conflicting with the corresponding local configuration.

DLDP implementation

- 1 On a DLDP-enabled link that is in up state, DLDP sends DLDP packets to the peer device and processes the DLDP packets received from the peer device. DLDP packets sent vary with DLDP states. Table 33 lists DLDP states and the corresponding packets.

Table 33 DLDP packet types and DLDP states

DLDP state	Type of DLDP packets sent
Active	Advertisement packet with RSY tag
Advertisement	Normal Advertisement packet
Probe	Probe packet
Disable	Disable packet and RecoverProbe packet



When a device transits from a DLDP state other than Inactive state or Disable state to Initial state, it sends Flush packets.

- 2 A received DLDP packet is processed as follows.
 - In any of the three authentication modes, the packet is dropped if it fails to pass the authentication.
 - The packet is dropped if the setting of the interval for sending Advertisement packets it carries conflicts with the corresponding local setting.
 - Other processes.

Table 34 Procedures for processing different types of DLDP packets

Packet type	Processing procedure
Advertisement packet with RSY tag	Retrieving the neighbor information. If the corresponding neighbor entry does not exist, creates the neighbor entry, triggers the Entry timer, and transits to Probe state. If the corresponding neighbor entry already exists, resets the Entry timer and transits to Probe state.
Normal Advertisement packet	Retrieves the neighbor information. If the corresponding neighbor entry does not exist, creates the neighbor entry, triggers the Entry timer, and transits to Probe state. If the corresponding neighbor entry already exists, resets the Entry timer.
Flush packet	Determines whether or not the local port is in Disable state. If yes, no process is performed. If not, removes the corresponding neighbor entry (if any).
Probe packet	Retrieves the neighbor information. If the corresponding neighbor entry does not exist, creates the neighbor entry, transits to Probe state, and returns Echo packets. If the corresponding neighbor entry already exists, resets the Entry timer and returns Echo packets.

Table 34 Procedures for processing different types of DLDP packets

Packet type	Processing procedure	
Echo packet	Retrieves the neighbor information.	<p>If the corresponding neighbor entry does not exist, creates the neighbor entry, triggers the Entry timer, and transits to Probe state.</p> <p>The corresponding neighbor entry already exists</p> <p>If the neighbor information it carries conflicts with the corresponding locally maintained neighbor entry, drops the packet.</p> <p>Otherwise, sets the flag of the neighbor as two-way connected. In addition, if the flags of all the neighbors are two-way connected, the device transits from Probe state to Advertisement state and disables the Echo timer.</p>
Disable packet	Check to see if the local port is in Disable state.	<p>If yes, no process is performed.</p> <p>If not, the local port transits to Disable state.</p>
RecoverProbe packet	Check to see if the local port is in Disable or Advertisement state.	<p>If not, no process is performed.</p> <p>If yes, returns RecoverEcho packets.</p>
RecoverEcho packet	Check to see if the local port is in Disable state.	<p>If not, no process is performed.</p> <p>If yes, the local port transits to Active state if the neighbor information the packet carries is consistent with the local port information.</p>
LinkDown packet	Check to see if the local port operates in Enhanced mode.	<p>If not, no process is performed.</p> <p>If yes and the local port is not in Disable state, the local transits to Disable state.</p>

- 3 If no echo packet is received from the neighbor, DLDP performs the following processing.

Table 35 Processing procedure when no echo packet is received from the neighbor

No echo packet received from the neighbor	Processing procedure
<p>In normal mode, no echo packet is received when the Echo timer expires.</p> <p>In enhanced mode, no echo packet is received when the enhanced timer expires.</p>	<p>DLDP transits to the Disable state, outputs log and tracking information, and sends Disable packets. In addition, depending on the user-defined DLDP down mode, DLDP shuts down the local port or prompts users to shut down the port, and removes the corresponding neighbor entry.</p>

DLDP neighbor state

A DLDP neighbor can be in one of the three states described in Table 36. You can check the state of a DLDP neighbor by using the **display dldp** command.

Table 36 Description on DLDP neighbor states

DLDP neighbor state	Description
Unknown	<p>A neighbor is in this state when it is just detected and is being probed. No information indicating the state of the neighbor is received. A neighbor is in this state only when it is being probed. It transits to Two way state or Unidirectional state after the probe operation finishes.</p>

Table 36 Description on DLDP neighbor states

DLDP neighbor state	Description
Two way	A neighbor is in this state after it receives response from its peer. This state indicates the link is a two-way link.
Unidirectional	A neighbor is in this state when the link connecting it is detected to be a unidirectional link. After a device transits to this state, the corresponding neighbor entries maintained on other devices are removed.

DLDP Configuration Task List

Complete the following tasks to configure DLDP:

Task	Remarks
“Enabling DLDP” on page 190	Required
“Setting DLDP Mode” on page 191	Optional
“Setting the Interval for Sending Advertisement Packets” on page 191	Optional
“Setting the DelayDown Timer” on page 191	Optional
“Setting the Port Shutdown Mode” on page 192	Optional
“Configuring DLDP Authentication” on page 192	Optional
“Resetting DLDP State” on page 193	Optional



- *DLDP works only when the link is up.*
- *To ensure unidirectional links can be detected, make sure these settings are the same on the both sides: DLDP state (enabled/disabled), the interval for sending Advertisement packets, authentication mode, and password.*
- *Keep the interval for sending Advertisement packets adequate to enable unidirectional links to be detected in time. If the interval is too long, unidirectional links cannot be terminated in time; if the interval is too short, network traffic may increase in vain.*
- *LACP (Link Aggregation Control Protocol) events have no effect on DLDP. Links in an aggregation group are treated individually in DLDP.*
- *802.1X has no effect on DLDP.*
- *When connecting two DLDP-enabled devices, make sure the DLDP version ID fields of the DLDP packets exchanged between the two devices are the same. Otherwise, DLDP may operate improperly.*

Enabling DLDP

Follow these steps to enable DLDP:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable DLDP globally	dldp enable	Required Globally disabled by default
Enter Ethernet port view or port group view	Enter Ethernet port view interface <i>interface-type interface-number</i> Enter port group view port-group { aggregation <i>agg-id</i> manual <i>port-group-name</i> }	Either of the two is required. The configuration performed in Ethernet port view applies to the current port only. The configuration performed in port group view applies to all the ports in the port group.

To do...	Use the command...	Remarks
Enable DLDP	dldp enable	Required Disabled on a port by default You can perform this operation on an optical port or an electrical port.



DLDP takes effect only when it is enabled both globally and on a port.

Setting DLDP Mode

Follow these steps to set DLDP mode:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Set DLDP mode	dldp work-mode { enhance normal }	Optional Normal by default

Setting the Interval for Sending Advertisement Packets

You can set the interval for sending Advertisement packets to enable unidirectional links to be detected in time.

Follow these steps to set the interval for sending Advertisement packets:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Set the interval for sending Advertisement packets	dldp interval time	Optional 5 seconds by default The interval for sending Advertisement packets applies to all the DLDP-enabled ports.



CAUTION:

- *Set the interval for sending Advertisement packets to a value not longer than one-third of the STP convergence time. If the interval is too long, STP loops may occur before unidirectional links are torn down; if the interval is too short, network traffic may increase in vain due to excessive Advertisement packets.*
- *To enable DLDP to operate properly, make sure the intervals for sending Advertisement packets on both sides of a link are the same.*

Setting the DelayDown Timer

On some ports, when the Tx line fails, the port goes down and then comes up again, causing optical signal jitters on the Rx line. When a port goes down due to a Tx failure, the device transits to the DelayDown state instead of the Inactive state to prevent the corresponding neighbor entries from being removed. In the same time, the device triggers the DelayDown timer. If the port goes up before the timer expires, the device restores the original state; if the port remains down when the timer expires, the devices transits to the Inactive state.

Follow these steps to set the DelayDown timer

To do...	Use the command...	Remarks
Enter system view	system-view	-
Set the DelayDown timer	dldp delaydown-timer <i>time</i>	Optional 1 second by default DelayDown timer setting applies to all the DLDP-enabled ports.

Setting the Port Shutdown Mode

On detecting a unidirectional link, the ports can be shut down in one of the following two modes.

- Manual mode. This mode applies to networks with low performance, where normal links may be treated as unidirectional links. It protects service packet transmission against false unidirectional links. In this mode, DLDP only detects unidirectional links and generates log and traps. The operations to shut down unidirectional link ports are accomplished by the administrator.
- Auto mode. In this mode, when a unidirectional link is detected, DLDP transits to Disable state, generates log and traps, and set the port as DLDP Down.

Follow these steps to set port shutdown mode

To do...	Use the command...	Remarks
Enter system view	system-view	-
Set port shutdown mode	dldp unidirectional-shutdown { auto manual }	Optional auto by default



CAUTION:

- *On a port with both remote OAM loopback and DLDP enabled, if the port shutdown mode is auto mode, the port will be shut down by DLDP when it receives a packet sent by itself, causing remote OAM loopback to operate improperly. To prevent this, you need to set the port shutdown mode to auto mode.*
- *If the device is busy, or the CPU utilization is high, normal links may be treated as unidirectional links. In this case, you can set the port shutdown mode to manual mode to eliminate the effects caused by false unidirectional link report.*

Configuring DLDP Authentication

Follow these steps to configure DLDP authentication:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure DLDP authentication	dldp authentication-mode { md5 <i>md5-password</i> none simple <i>simple-password</i> }	Required none by default



CAUTION: *To enable DLDP to operate properly, make sure the DLDP authentication modes and the passwords of the both sides of a link are the same.*

Resetting DLDP State

After a unidirectional link is detected, DLDP shuts down the corresponding port. To enable the port to perform DLDP detect again, you can reset DLDP state for it. A port can be in different state after you reset DLDP state for it. That is, it can be in Inactive state (if the port is physically down) or in Active state (if the port is physically up) after you reset DLDP state for it.



CAUTION:

- The configuration of resetting DLDP state performed in system view applies to all the ports shut down by DLDP.
- The configuration of resetting DLDP state performed in port view or port group view applies to the current port or all the ports in the port group shut down by DLDP.

Resetting DLDP State in System view

Follow these steps to reset DLDP in system view:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Reset DLDP state	dldp reset	Required

Resetting DLDP State in Port view/Port Group View

Follow these steps to reset DLDP state in port view/port group view:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view/port group view	Enter Ethernet port view interface <i>interface-type interface-number</i>	Either is required. The configuration performed in Ethernet port view applies to the current port only; the configuration performed in port group view applies to all the ports in the port group.
Reset DLDP state	port-group { aggregation <i>agg-id</i> manual <i>port-group-name</i> }	Required
Reset DLDP state	dldp reset	Required

Displaying and Maintaining DLDP

To do...	Use the command...	Remarks
Display the DLDP configuration of a port	display dldp [<i>interface-type interface-number</i>]	Available in any view
Display the statistics on DLDP packets passing through a port	display dldp statistics [<i>interface-type interface-number</i>]	Available in any view
Clear the statistics on DLDP packets passing through a port	reset dldp statistics [<i>interface-type interface-number</i>]	Available in user view

DLDP Configuration Example

DLDP Configuration Example

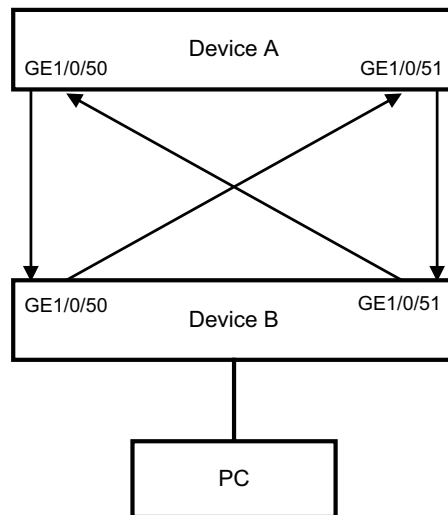
Network requirements

- Device A and Device B are connected through two fiber pairs, in which two fibers are cross-connected, as shown in Figure 52.

- It is desired that the unidirectional links can be disconnected on being detected; and the ports shut down by DLDP can be restored after the fiber connections are corrected.

Network diagram

Figure 52 Network diagram for DLDP configuration



Configuration procedure

1 Configuration on Device A

Enable DLDP on GigabitEthernet1/0/50 and GigabitEthernet 1/0/51.

```

<DeviceA> system-view
[DeviceA] interface gigabitEthernet 1/0/50
[DeviceA-GigabitEthernet1/0/50] dldp enable
[DeviceA-GigabitEthernet1/0/50] interface gigabitEthernet 1/0/51
[DeviceA-GigabitEthernet1/0/51] dldp enable
[DeviceA-GigabitEthernet1/0/51] quit
  
```

Set the interval for sending Advertisement packets to 6 seconds.

```
[DeviceA] dldp interval 6
```

Set the DelayDown timer to 2 seconds.

```
[DeviceA] dldp delaydown-timer 2
```

Set the DLDP mode as enhanced mode.

```
[DeviceA] dldp work-mode enhance
```

Set the port shutdown mode as auto mode.

```
[DeviceA] dldp unidirectional-shutdown auto
```

Enable DLDP globally.

```
[DeviceA] dldp enable
```

Check the information about DLDAP.

```
[DeviceA] display dldp
DLDP global status : enable
DLDP interval : 6s
DLDP work-mode : enhance
DLDP authentication-mode : none
DLDP unidirectional-shutdown : auto
DLDP delaydown-timer : 2s
The number of enabled ports is 2.
```

```
Interface GigabitEthernet1/0/50
DLDP port state : disable
DLDP link state : down
The neighbor number of the port is 0.
```

```
Interface GigabitEthernet1/0/51
DLDP port state : disable
DLDP link state : down
The neighbor number of the port is 0.
```

The output information indicates that both GigabitEthernet1/0/50 and GigabitEthernet1/0/51 are in Disable state and the links are down, which means unidirectional links are detected and the two ports are thus shut down.

Reset DLDAP state for the ports shut down by DLDAP.

```
[DeviceA] dldp reset
```

2 Configuration on Device B

The configuration on Device B is the same as that on Device A and is thus omitted.



If two fibers are cross-connected, all the four ports involved will be shut down by DLDAP.

Troubleshooting

Symptom:

Two DLDAP-enabled devices, Device A and Device B, are connected through two fiber pairs, in which two fibers are cross-connected. The unidirectional links cannot be detected; all the four ports involved are in Advertisement state.

Analysis:

The problem can be caused by the following.

- The intervals for sending Advertisement packets on Device A and Device B are not the same.
- DLDAP authentication modes/passwords on Device A and Device B are not the same.

Solution:

Make sure the interval for sending Advertisement packets, the authentication mode, and the password on Device A and Device B are the same.

When configuring MSTP, go to these sections for information you are interested in:

- “MSTP Overview” on page 197
- “Configuring the Root Bridge” on page 213
- “Configuring Leaf Nodes” on page 224
- “Performing mCheck” on page 228
- “Configuring Protection Functions” on page 233
- “Displaying and Maintaining MSTP” on page 235

MSTP Overview

Introduction to STP **Why STP?**

The Spanning Tree Protocol (STP) was established based on the 802.1d standard of IEEE to eliminate physical loops at the data link layer in a local area network (LAN). Devices running this protocol detect loops in the network by exchanging information with one another and eliminate loops by selectively blocking certain ports until the loop structure is pruned into a loop-free network structure. This avoids proliferation and infinite recycling of packets that would occur in a loop network and prevents deterioration of the packet processing capability of network devices caused by duplicate packets received.

In the narrow sense, STP refers to the STP protocol defined in IEEE 802.1d; in the broad sense, it refers to the STP protocol defined in IEEE 802.1d and various enhanced spanning tree protocols derived from the STP protocol.

Protocol Packets of STP

STP uses bridge protocol data units (BPDUs), also known as configuration messages, as its protocol packets.

STP identifies the network topology by transmitting BPDUs between STP-compliant network devices. BPDUs contain sufficient information for the network devices to complete the spanning tree calculation.

In STP, BPDUs come in two types:

- Configuration BPDUs, used for calculating spanning trees and maintaining the spanning tree topology.
- Topology change notification (TCN) BPDUs, used for notifying concerned devices of network topology changes, if any.

Basic concepts in STP

1 Root bridge

A tree network must have a root; hence the concept of “root bridge” has been introduced in STP.

There is one and only one root bridge in the entire network, and the root bridge can change along with changes of the network topology. Therefore, the root bridge is not fixed.

Upon network convergence, the root bridge generates and sends out configuration BPDUs at a certain interval, and other devices just forward the BPDUs. This mechanism ensures topological stability.

2 Root port

On a non-root bridge device, the root port is the port nearest to the root bridge. The root port is responsible for communication with the root bridge. A non-root-bridge device has one and only one root port. The root bridge has no root port.

3 Designated bridge and designated port

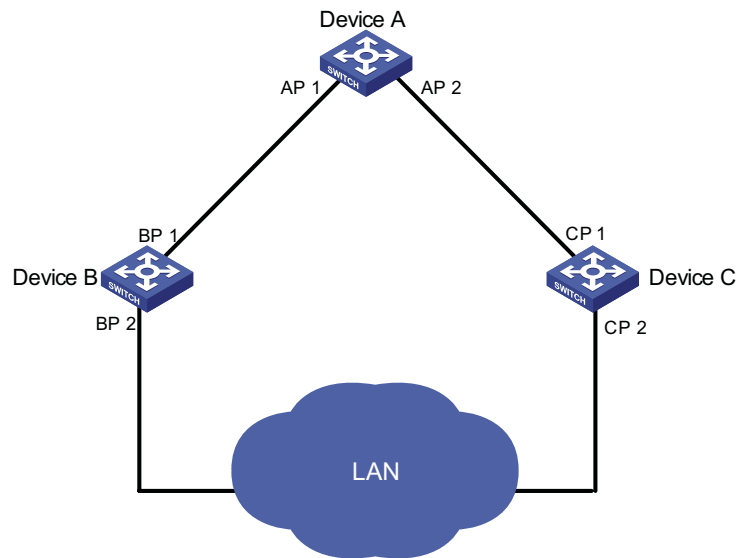
The following table describes a designated bridge and a designated port.

Table 37 Description of designated bridge and designated port

Classification	Designated bridge	Designated port
For a device	The device directly connected with this device and responsible for forwarding BPDUs	The port through which the designated bridge forwards BPDUs to this device
For a LAN	The device responsible for forwarding BPDUs to this LAN segment	The port through which the designated bridge forwards BPDUs to this LAN segment

Figure 53 shows designated bridges and designated ports. In the figure, AP1 and AP2, BP1 and BP2, and CP1 and CP2 are ports on Device A, Device B, and Device C respectively.

- If Device A forwards BPDUs to Device B through AP1, the designated bridge for Device B is Device A, and the designated port is the port AP1 on Device A.
- Two devices are connected to the LAN: Device B and Device C. If Device B forwards BPDUs to the LAN, the designated bridge for the LAN is Device B, and the designated port is the port BP2 on Device B.

Figure 53 A schematic diagram of designated bridges and designated ports**Path cost**

Path cost is a reference value used for link selection in STP. By calculating the path cost, STP selects relatively "robust" links and blocks redundant links, and finally prunes the network into loop-free tree structure.



All the ports on the root bridge are designated ports.

How STP works

STP identifies the network topology by transmitting configuration BPDUs between network devices. Configuration BPDUs contain sufficient information for network devices to complete the spanning tree calculation. Important fields in a configuration BPDU include:

- Root bridge ID: consisting of root bridge priority and MAC address.
- Root path cost: the cost of the shortest path to the root bridge.
- Designated bridge ID: designated bridge priority plus MAC address.
- Designated port ID, designated port priority plus port name.
- Message age: age of the configuration BPDU while it propagates in the network.
- Max age: maximum age of the configuration BPDU maintained in the device.
- Hello time: configuration BPDU interval.
- Forward delay: forward delay of the port.



For the convenience of description, the description and examples below involve only four parts of a configuration BPDU

- *Root bridge ID (in the form of device priority)*
- *Root path cost*
- *Designated bridge ID (in the form of device priority)*

- Designated port ID (in the form of port name)
- 1 Specific calculation process of the STP algorithm

- Initial state

Upon initialization of a device, each port generates a BPDU with itself as the root bridge, in which the root path cost is 0, designated bridge ID is the device ID, and the designated port is the local port.

- Selection of the optimum configuration BPDU

Each device sends out its configuration BPDUs and receives configuration BPDUs from other devices.

The process of selecting the optimum configuration BPDU is as follows:

Table 38 Selection of the optimum configuration BPDU

Step	Description
1	<p>Upon receiving a configuration BPDU on a port, the device performs the following processing:</p> <ul style="list-style-type: none"> ■ If the received configuration BPDU has a lower priority than that of the configuration BPDU generated by the port, the device will discard the received configuration BPDU without doing any processing on the configuration BPDU of this port. ■ If the received configuration BPDU has a higher priority than that of the configuration BPDU generated by the port, the device will replace the content of the configuration BPDU generated by the port with the content of the received configuration BPDU.
2	The device compares the configuration BPDUs of all the ports and chooses the optimum configuration BPDU.



Principle for configuration BPDU comparison

- *The configuration BPDU that has the lowest root bridge ID has the highest priority.*
- *If all the configuration BPDUs have the same root bridge ID, they will be compared for their root path costs. If the root path cost in a configuration BPDU plus the path cost corresponding to this port is S , the configuration BPDU with the smallest S value has the highest priority.*
- *If all configuration BPDUs have the same root path cost, they will be compared for their designated bridge IDs, then their designated port IDs, and then the IDs of the ports on which they are received. The smaller the ID, the higher message priority.*
- *Selection of the root bridge*

At network initialization, each STP-compliant device on the network assumes itself to be the root bridge, with the root bridge ID being its own device ID. By exchanging configuration BPDUs, the devices compare one another's root bridge ID. The device with the smallest root bridge ID is elected as the root bridge.

- Selection of the root port and designated ports

The process of selecting the root port and designated ports is as follows:

Table 39 Selection of the root port and designated ports

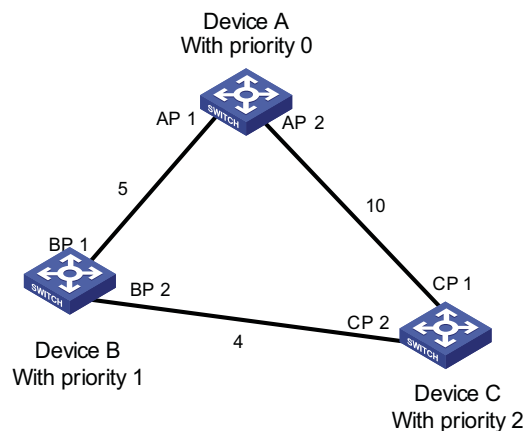
Step	Description
1	A non-root-ridge device regards the port on which it received the optimum configuration BPDU as the root port.
2	Based on the configuration BPDU and the path cost of the root port, the device calculates a designated port configuration BPDU for each of the rest ports. <ul style="list-style-type: none"> ■ The root bridge ID is replaced with that of the configuration BPDU of the root port. ■ The root path cost is replaced with that of the configuration BPDU of the root port plus the path cost corresponding to the root port. ■ The designated bridge ID is replaced with the ID of this device. ■ The designated port ID is replaced with the ID of this port.
3	The device compares the calculated configuration BPDU with the configuration BPDU on the port of which the port role is to be defined, and does different things according to the comparison result: <ul style="list-style-type: none"> ■ If the calculated configuration BPDU is superior, the device will consider this port as the designated port, and the configuration BPDU on the port will be replaced with the calculated configuration BPDU, which will be sent out periodically. ■ If the configuration BPDU on the port is superior, the device will block this port without updating its configuration BPDU, so that the port will only receive BPDUs, but not send any, and will not forward data.



When the network topology is stable, only the root port and designated ports forward traffic, while other ports are all in the blocked state - they only receive STP packets but do not forward user traffic.

Once the root bridge, the root port on each non-root bridge and designated ports have been successfully elected, the entire tree-shaped topology has been constructed.

The following is an example of how the STP algorithm works. The specific network diagram is shown in Figure 54. In the feature, the priority of Device A is 0, the priority of Device B is 1, the priority of Device C is 2, and the path costs of these links are 5, 10 and 4 respectively.

Figure 54 Network diagram for the STP algorithm

- Initial state of each device

The following table shows the initial state of each device.

Table 40 Initial state of each device

Device	Port name	BPDU of port
Device A	AP1	{0, 0, 0, AP1}
	AP2	{0, 0, 0, AP2}
Device B	BP1	{1, 0, 1, BP1}
	BP2	{1, 0, 1, BP2}
Device C	CP1	{2, 0, 2, CP1}
	CP2	{2, 0, 2, CP2}

- Comparison process and result on each device

The following table shows the comparison process and result on each device.

Table 41 Comparison process and result on each device

Device	Comparison process	BPDU of port after comparison
Device A	<ul style="list-style-type: none"> ■ Port AP1 receives the configuration BPDU of Device B {1, 0, 1, BP1}. Device A finds that the configuration BPDU of the local port {0, 0, 0, AP1} is superior to the configuration received message, and discards the received configuration BPDU. ■ Port AP2 receives the configuration BPDU of Device C {2, 0, 2, CP1}. Device A finds that the BPDU of the local port {0, 0, 0, AP2} is superior to the received configuration BPDU, and discards the received configuration BPDU. ■ Device A finds that both the root bridge and designated bridge in the configuration BPDUs of all its ports are Device A itself, so it assumes itself to be the root bridge. In this case, it does not make any change to the configuration BPDU of each port, and starts sending out configuration BPDUs periodically. 	AP1: {0, 0, 0, AP1} AP2: {0, 0, 0, AP2}

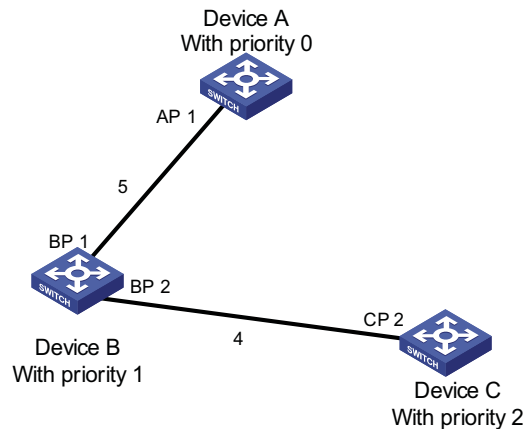
Table 41 Comparison process and result on each device

Device	Comparison process	BPDU of port after comparison
Device B	<ul style="list-style-type: none"> <li data-bbox="662 352 1295 485">■ Port BP1 receives the configuration BPDU of Device A {0, 0, 0, AP1}. Device B finds that the received configuration BPDU is superior to the configuration BPDU of the local port {1, 0, 1, BP1}, and updates the configuration BPDU of BP1. <li data-bbox="662 499 1295 632">■ Port BP2 receives the configuration BPDU of Device C {2, 0, 2, CP2}. Device B finds that the configuration BPDU of the local port {1, 0, 1, BP2} is superior to the received configuration BPDU, and discards the received configuration BPDU. <li data-bbox="662 646 1295 779">■ Device B compares the configuration BPDUs of all its ports, and determines that the configuration BPDU of BP1 is the optimum configuration BPDU. Then, it uses BP1 as the root port, the configuration BPDUs of which will not be changed. <li data-bbox="662 793 1295 869">■ Based on the configuration BPDU of BP1 and the path cost of the root port (5), Device B calculates a designated port configuration BPDU for BP2 {0, 5, 1, BP2}. <li data-bbox="662 884 1295 1043">■ Device B compares the calculated configuration BPDU {0, 5, 1, BP2} with the configuration BPDU of BP2. If the calculated BPDU is superior, BP2 will act as the designated port, and the configuration BPDU on this port will be replaced with the calculated configuration BPDU, which will be sent out periodically. 	<p data-bbox="1312 352 1451 407">BP1: {0, 0, 0, AP1}</p> <p data-bbox="1312 422 1451 476">BP2: {1, 0, 1, BP2}</p> <p data-bbox="1312 646 1468 722">Root port BP1: {0, 0, 0, AP1}</p> <p data-bbox="1312 728 1484 783">Designated port BP2: {0, 5, 1, BP2}</p>

Table 41 Comparison process and result on each device

Device	Comparison process	BPDU of port after comparison
Device C	<ul style="list-style-type: none"> Port CP1 receives the configuration BPDU of Device A {0, 0, 0, AP2}. Device C finds that the received configuration BPDU is superior to the configuration BPDU of the local port {2, 0, 2, CP1}, and updates the configuration BPDU of CP1. Port CP2 receives the configuration BPDU of port BP2 of Device B {1, 0, 1, BP2} before the message was updated. Device C finds that the received configuration BPDU is superior to the configuration BPDU of the local port {2, 0, 2, CP2}, and updates the configuration BPDU of CP2. 	CP1: {0, 0, 0, AP2} CP2: {1, 0, 1, BP2}
	By comparison: <ul style="list-style-type: none"> The configuration BPDU of CP1 is elected as the optimum configuration BPDU, so CP1 is identified as the root port, the configuration BPDUs of which will not be changed. Device C compares the calculated designated port configuration BPDU {0, 10, 2, CP2} with the configuration BPDU of CP2, and CP2 becomes the designated port, and the configuration BPDU of this port will be replaced with the calculated configuration BPDU. Next, port CP2 receives the updated configuration BPDU of Device B {0, 5, 1, BP2}. Because the received configuration BPDU is superior to its old one, Device C launches a BPDU update process. At the same time, port CP1 receives configuration BPDUs periodically from Device A. Device C does not launch an update process after comparison. 	Root port CP1: {0, 0, 0, AP2} Designated port CP2: {0, 10, 2, CP2}
	By comparison: <ul style="list-style-type: none"> Because the root path cost of CP2 (9) (root path cost of the BPDU (5) plus path cost corresponding to CP2 (4)) is smaller than the root path cost of CP1 (10) (root path cost of the BPDU (0) + path cost corresponding to CP2 (10)), the BPDU of CP2 is elected as the optimum BPDU, and CP2 is elected as the root port, the messages of which will not be changed. After comparison between the configuration BPDU of CP1 and the calculated designated port configuration BPDU, port CP1 is blocked, with the configuration BPDU of the port remaining unchanged, and the port will not receive data from Device A until a spanning tree calculation process is triggered by a new condition, for example, the link from Device B to Device C becomes down. 	CP1: {0, 0, 0, AP2} CP2: {0, 5, 1, BP2}
		Blocked port CP2: {0, 0, 0, AP2} Root port CP2: {0, 5, 1, BP2}

After the comparison processes described in the table above, a spanning tree with Device A as the root bridge is stabilized, as shown in Figure 55.

Figure 55 The final calculated spanning tree

To facilitate description, the spanning tree calculation process in this example is simplified, while the actual process is more complicated.

2 The BPDU forwarding mechanism in STP

- Upon network initiation, every switch regards itself as the root bridge, generates configuration BPDUs with itself as the root, and sends the configuration BPDUs at a regular interval of hello time.
- If it is the root port that received the configuration BPDU and the received configuration BPDU is superior to the configuration BPDU of the port, the device will increase message age carried in the configuration BPDU by a certain rule and start a timer to time the configuration BPDU while it sends out this configuration BPDU through the designated port.
- If the configuration BPDU received on the designated port has a lower priority than the configuration BPDU of the local port, the port will immediately send out its better configuration BPDU in response.
- If a path becomes faulty, the root port on this path will no longer receive new configuration BPDUs and the old configuration BPDUs will be discarded due to timeout. In this case, the device will generate a configuration BPDU with itself as the root and sends out the BPDU. This triggers a new spanning tree calculation process so that a new path is established to restore the network connectivity.

However, the newly calculated configuration BPDU will not be propagated throughout the network immediately, so the old root ports and designated ports that have not detected the topology change continue forwarding data along the old path. If the new root port and designated port begin to forward data as soon as they are elected, a temporary loop may occur.

3 STP timers

STP calculations need three important timing parameters: forward delay, hello time, and max age.

- Forward delay is the delay time for device state transition. A path failure will cause re-calculation of the spanning tree, and the spanning tree structure will change accordingly. However, the new configuration BPDU as the calculation result cannot be propagated throughout the network immediately. If the newly elected root port and designated ports start to forward data right away, a temporary loop is likely to occur. For this reason, as a mechanism for state

transition in STP, a newly elected root port or designated port requires twice the forward delay time before transitioning to the forwarding state, when the new configuration BPDU has been propagated throughout the network.

- Hello time is the time interval at which a device sends hello packets to the surrounding devices to make sure that the paths are fault-free.
- Max age is a parameter used to determine whether a configuration BPDU held in the device has expired. A configuration BPDU beyond the max age will be discarded.

Introduction to MSTP **Why MSTP**

1 Disadvantages of STP and RSTP

STP does not support rapid state transition of ports. A newly elected root port or designated port must wait twice the forward delay time before transitioning to the forwarding state, even if it is a port on a point-to-point link or it is an edge port, which directly connects to a user terminal rather than to another device or a shared LAN segment.

The rapid spanning tree protocol (RSTP) is an optimized version of STP. RSTP allows a newly elected root port or designated port to enter the forwarding state much quicker under certain conditions than in STP. As a result, it takes a shorter time for the network to reach the final topology stability.



In RSTP, a newly elected root port can enter the forwarding state rapidly if this condition is met: The old root port on the device has stopped forwarding data and the upstream designated port has started forwarding data.

In RSTP, a newly elected designated port can enter the forwarding state rapidly if this condition is met: The designated port is an edge port or a port connected with a point-to-point link. If the designated port is an edge port, it can enter the forwarding state directly; if the designated port is connected with a point-to-point link, it can enter the forwarding state immediately after the device undergoes handshake with the downstream device and gets a response.

Although RSTP support rapid network convergence, it has the same drawback as STP does: All bridges within a LAN share the same spanning tree, so redundant links cannot be blocked based on VLANs, and the packets of all VLANs are forwarded along the same spanning tree.

2 Features of MSTP

The multiple spanning tree protocol (MSTP) overcomes the shortcomings of STP and RSTP. In addition to support for rapid network convergence, it also allows data flows of different VLANs to be forwarded along their own paths, thus providing a better load sharing mechanism for redundant links. For description about VLANs, refer to “Introduction to VLAN” on page 83.

MSTP features the following:

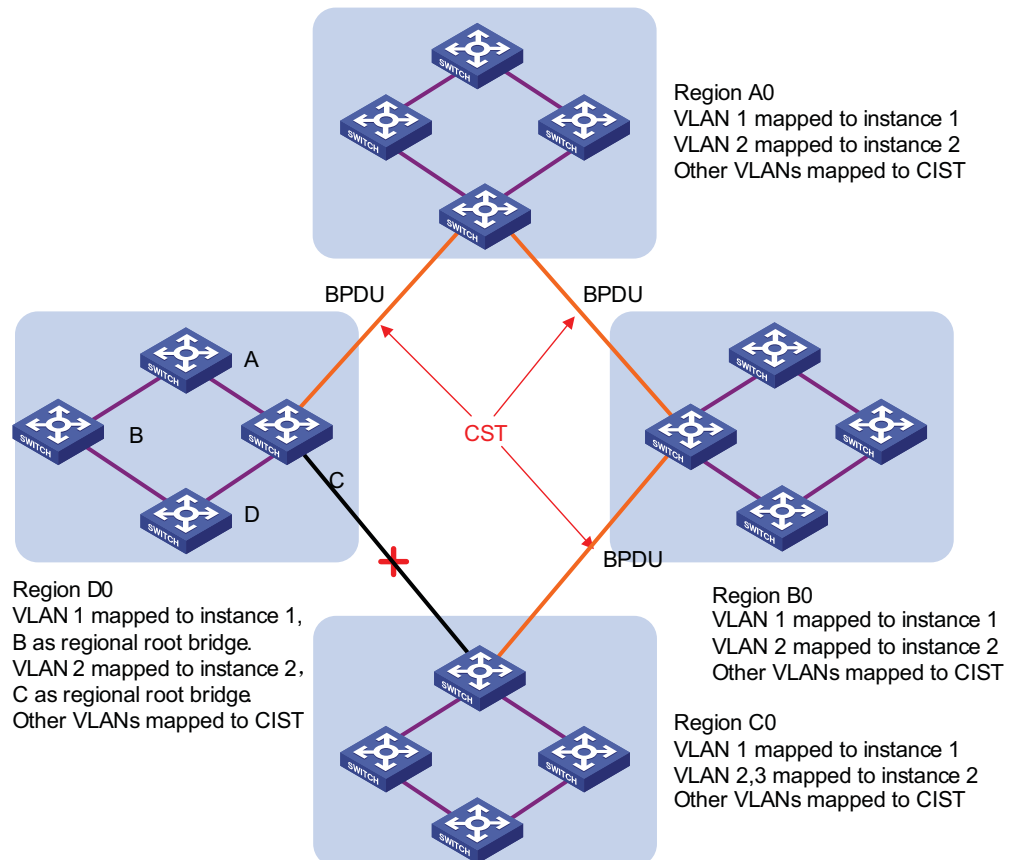
- MSTP supports mapping VLANs to MST instances by means of a VLAN-to-instance mapping table. MSTP can save communication overheads and resource usage by mapping multiple VLANs to one MST instance.
- MSTP divides a switched network into multiple regions, each containing multiple spanning trees that are independent of one another.

- MSTP prunes loop networks into a loop-free tree, thus avoiding proliferation and endless recycling of packets in a loop network. In addition, it provides multiple redundant paths for data forwarding, thus supporting load balancing of VLAN data in the data forwarding process.
- MSTP is compatible with STP and RSTP.

Basic concepts in MSTP

Assume that all the four switches in Figure 56 are running MSTP. In light with the diagram, the following paragraphs will present some basic concepts of MSTP.

Figure 56 Basic concepts in MSTP



1 MST region

A multiple spanning tree region (MST region) is composed of multiple devices in a switched network and network segments among them. These devices have the following characteristics:

- All are MSTP-enabled,
- They have the same region name,
- They have the same VLAN-to-instance mapping configuration,
- They have the same MSTP revision level configuration, and
- They are physically linked with one another.

For example, all the devices in region A0 in Figure 56 have the same MST region configuration:

- The same region name,
- The same VLAN-to-instance mapping (VLAN 1 is mapped to MST instance 1, VLAN 2 to MST instance 2, and the rest to the command and internal spanning tree (CIST). CIST refers to MST instance 0), and
- The same MSTP revision level (not shown in the figure).

Multiple MST regions can exist in a switched network. You can use an MSTP command to group multiple devices to the same MST region.

2 VLAN-to-instance mapping table

As an attribute of an MST region, the VLAN-to-instance mapping table describes the mapping relationships between VLANs and MST instances. In Figure 56, for example, the VLAN-to-instance mapping table of region A0 describes that the same region name, the same VLAN-to-instance mapping (VLAN 1 is mapped to MST instance 1, VLAN 2 to MST instance 2, and the rest to CIST). MSTP achieves load balancing by means of the VLAN-to-instance mapping table.

3 IST

Internal spanning tree (IST) is a spanning tree that runs in an MST region.

ISTs in all MST regions and the common spanning tree (CST) jointly constitute the common and internal spanning tree (CIST) of the entire network. An IST is a section of the CIST in the given MST region.

In Figure 56, for example, the CIST has a section in each MST region, and this section is the IST in the respective MST region.

4 CST

The CST is a single spanning tree that connects all MST regions in a switched network. If you regard each MST region as a “device”, the CST is a spanning tree calculated by these devices through STP or RSTP. For example, the red lines in Figure 56 describe the CST.

5 CIST

Jointly constituted by ISTs and the CST, the CIST is a single spanning tree that connects all devices in a switched network.

In Figure 56, for example, the ISTs in all MST regions plus the inter-region CST constitute the CIST of the entire network.

6 MSTI

Multiple spanning trees can be generated in an MST region through MSTP, one spanning tree being independent of another. Each spanning tree is referred to as a multiple spanning tree instance (MSTI). In Figure 56, for example, multiple spanning tree can exist in each MST region, each spanning tree corresponding to a VLAN. These spanning trees are called MSTIs.

7 Regional root bridge

The root bridge of the IST or an MSTI within an MST region is the regional root bridge of the MST or that MSTI. Based on the topology, different spanning trees in an MST region may have different regional roots.

For example, in region D0 in Figure 56, the regional root of instance 1 is device B, while that of instance 2 is device C.

8 Common root bridge

The common root bridge is the root bridge of the CIST.

In Figure 56, for example, the common root bridge is a device in region A0.

9 Boundary port

A boundary port is a port that connects an MST region to another MST configuration, or to a single spanning-tree region running STP, or to a single spanning-tree region running RSTP.

During MSTP calculation, a boundary port assumes the same role on the CIST and on MST instances. Namely, if a boundary port is the master port on the CIST, it is also the master port on all MST instances within this region. In Figure 56, for example, if a device in region A0 is interconnected with the first port of a device in region D0 and the common root bridge of the entire switched network is located in region A0, the first port of that device in region D0 is the boundary port of region D0.



Currently, the device is not capable of recognizing boundary ports. When the device interworks with a third party's device that supports boundary port recognition, the third party's device may malfunction in recognizing a boundary port.

10 Roles of ports

In the MSTP calculation process, port roles include root port, designated port, master port, alternate port, backup port, and so on.

- Root port: a port responsible for forwarding data to the root bridge.
- Designated port: a port responsible for forwarding data to the downstream network segment or device.
- A master port connects an MST region to the common root. The path from the master port to the common root is the shortest path between the MST region and the common root. In the CST, the master port is the root port of the region, which is considered as a node. The master port is a special boundary port. It is a root port in the IST/CIST while a master port in the other MSTIs.
- Alternate port: The standby port for the root port or master port. When the root port or master port is blocked, the alternate port becomes the new root port or master port.
- Backup port: The backup port of designated ports. When a designated port is blocked, the backup port becomes a new designated port and starts forwarding data without delay. When a loop occurs while two ports of the same MSTP device are interconnected, the device will block either of the two ports, and the backup port is that port to be blocked.

A port can assume different roles in different MST instances.

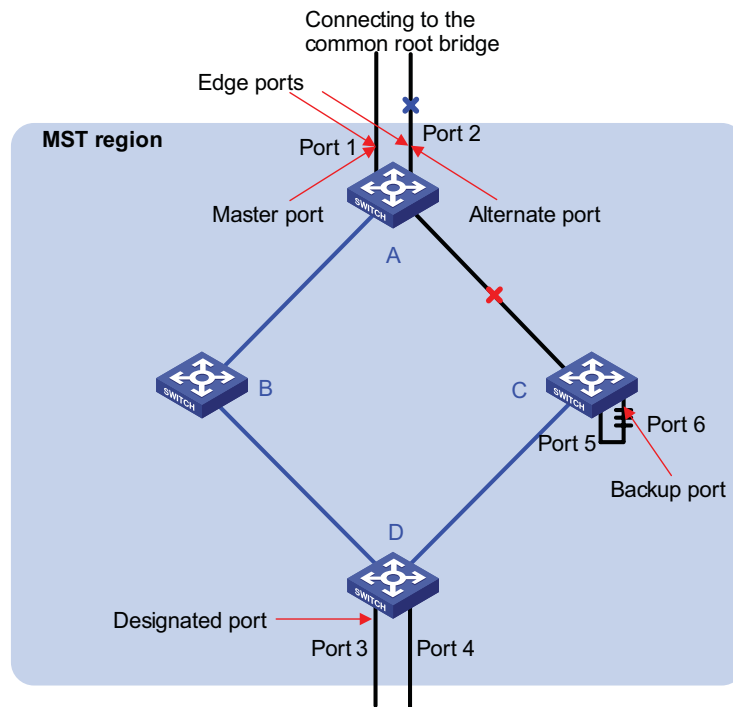
Figure 57 Port roles

Figure 57 helps understand these concepts. Where,

- Devices A, B, C, and D constitute an MST region.
- Port 1 and port 2 of device A connect to the common root bridge.
- Port 5 and port 6 of device C form a loop.
- Port 3 and port 4 of device D connect downstream to other MST regions.

11 Port states

In MSTP, port states fall into the following tree:

- Forwarding: the port learns MAC addresses and forwards user traffic;
- Learning: the port learns MAC addresses but does not forward user traffic;
- Discarding: the port neither learns MAC addresses nor forwards user traffic.



When in different MST instances, a port can be in different states.

- The role a boundary port plays in an MSTI is consistent with the role it plays in the CIST. The master port, which is a root port in the CIST while a master port in the other MSTIs, is an exception.
- For example, in Figure 57, port 1 on switch A is a boundary port. It is a root port in the CIST while a master port in all the other MSTIs in the region.

A port state is not exclusively associated with a port role. Table 42 lists the port state(s) supported by each port role ("," indicates that the port supports this state, while "-" indicates that the port does not support this state).

Table 42 Ports states supported by different port roles

Role/ State	Root port/ Master port	Designated port	Alternate port	Backup port
Forwarding	Yes	Yes	No	No
Learning	Yes	Yes	No	No
Discarding	Yes	Yes	Yes	Yes

How MSTP works

MSTP divides an entire Layer 2 network into multiple MST regions, which are interconnected by a calculated CST. Inside an MST region, multiple spanning trees are generated through calculation, each spanning tree called an MST instance. Among these MST instances, instance 0 is the IST, while all the others are MSTIs. Similar to STP, MSTP uses configuration BPDUs to calculate spanning trees. The only difference between the two protocols is that an MSTP BPDU carries the MSTP configuration on the device from which this BPDU is sent.

1 CIST calculation

By comparison of configuration BPDUs, the device with the highest priority is elected as the root bridge of the CIST. MSTP generates an IST within each MST region through calculation, and, at the same time, MSTP regards each MST region as a single device and generates a CST among these MST regions through calculation. The CST and ISTs constitute the CIST of the entire network.

2 MSTI calculation

Within an MST region, MSTP generates different spanning tree instances for different VLANs based on the VLAN-to-instance mappings. MSTP performs a separate calculation process, which is similar to spanning tree calculation in STP, for each spanning tree. For details, refer to “How STP works” on page 199.

In MSTP, a VLAN packet is forwarded along the following paths:

- Within an MST region, the packet is forwarded along the corresponding MSTI.
- Between two MST regions, the packet is forwarded along the CST.

Implementation of MSTP on devices

MSTP is compatible with STP and RSTP. STP and RSTP protocol packets can be recognized by devices running MSTP and used for spanning tree calculation.

In addition to basic MSTP functions, many management-facilitating special functions are provided, as follows:

- Root bridge hold
- Root bridge backup
- Root guard
- BPDU guard
- Loop guard
- TC-BPDU guard

Protocols and Standards MSTP is documented in:

- IEEE 802.1d: Spanning Tree Protocol

- IEEE 802.1w: Rapid Spanning Tree Protocol
- IEEE 802.1s: Multiple Spanning Tree Protocol

Configuration Task List

Before configuring MSTP, you need to know the position of each device in each MST instance: root bridge or leaf node. In each instance, one, and only one device acts as the root bridge, while all others as leaf nodes.

Complete these tasks to configure MSTP:

Task	Remarks
"Configuring the Root Bridge" on page 213	Required
"Configuring an MST Region" on page 213	Optional
"Specifying the Root Bridge or a Secondary Root Bridge" on page 215	Optional
"Configuring the Work Mode of MSTP Device" on page 216	Optional
"Configuring the Priority of the Current Device" on page 216	Optional
"Configuring the Maximum Hops of an MST Region" on page 217	Optional
"Configuring the Network Diameter of a Switched Network" on page 218	Optional
"Configuring Timers of MSTP" on page 218	Optional
"Configuring the Timeout Factor" on page 219	Optional
"Configuring the Maximum Transmission Rate of Ports" on page 220	Optional
"Configuring Ports as Edge Ports" on page 221	Optional
"Configuring Whether Ports Connect to Point-to-Point Links" on page 221	Optional
"Configuring the Mode a Port Uses to Recognize/Send MSTP Packets" on page 222	Optional
"Enabling the Output of Port State Transition Information" on page 223	Optional
"Enabling the MSTP Feature" on page 224	Required

Task	Remarks
"Configuring Leaf Nodes" on page 224	"Configuring an MST Region" on page 213 Required
	"Configuring the Work Mode of MSTP Device" on page 216 Optional
	"Configuring the Timeout Factor" on page 219 Optional
	"Configuring the Maximum Transmission Rate of Ports" on page 220 Optional
	"Configuring Ports as Edge Ports" on page 221 Optional
	"Configuring Path Costs of Ports" on page 225 Optional
	"Configuring Port Priority" on page 226 Optional
	"Configuring Whether Ports Connect to Point-to-Point Links" on page 221 Optional
	"Configuring the Mode a Port Uses to Recognize/Send MSTP Packets" on page 227 Optional
	"Enabling the Output of Port State Transition Information" on page 223 Optional
	"Enabling the MSTP Feature" on page 224 Required
"Performing mCheck" on page 228	Optional
"Configuring Digest Snooping" on page 229	Optional
"Configuring No Agreement Check" on page 230	Optional
"Configuring Protection Functions" on page 233	Optional



In a network containing switches with both GVRP and MSTP enabled, GVRP messages travel along the CIST. If you want to advertise a VLAN through GVRP, be sure to map the VLAN to the CIST (MSTI 0) when configuring the VLAN-to-MSTI mapping table. For detailed information of GVRP, refer to "GVRP Configuration" on page 109.

Configuring the Root Bridge

Configuring an MST Region

Configuration procedure

Follow these steps to configure an MST region:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter MST region view	stp region-configuration	-
Configure the MST region name	region-name <i>name</i>	Optional The MST region name is the MAC address by default.

To do...	Use the command...	Remarks
Configure the VLAN-to-instance mapping table	instance <i>instance-id</i> vlan <i>vlan-list</i> vlan-mapping modulo <i>modulo</i>	Optional Use either command. All VLANs in an MST region are mapped to MST instance 0 by default.
Configure the MSTP revision level of the MST region	revision-level <i>level</i>	Optional 0 by default
Activate MST region configuration manually	active region-configuration	Required
Display all the configuration information of the MST region	check region-configuration	Optional
Display the currently effective MST region configuration information	display stp region-configuration	The display command can be executed in any view.



- *MSTP-enabled switches are in the same region only when they have the same format selector (a 802.1s-defined protocol selector, which is 0 by default and cannot be configured), MST region name, VLAN-to-MSTI mapping table, and revision level.*
- *The 3Com series support only the MST region name, VLAN-to-MSTI mapping table, and revision level. Switches with the settings of these parameters being the same are assigned to the same MST region.*

The configuration of MST region-related parameters, especially the VLAN-to-instance mapping table, will cause MSTP to launch a new spanning tree calculation process, which may result in network topology instability. To reduce the possibility of topology instability caused by configuration, MSTP will not immediately launch a new spanning tree calculation process when processing MST region-related configurations; instead, such configurations will take effect only if you:

- activate the MST region-related parameters using the **active region-configuration** command, or
- enable MSTP using the **stp enable** command.

Configuration example

Configure the MST region name to be "info", the MSTP revision level to be 1, and VLAN 2 through VLAN 10 to be mapped to instance 1 and VLAN 20 through VLAN 30 to instance 2.

```
<Sysname> system-view
[Sysname] stp region-configuration
[Sysname-mst-region] region-name info
[Sysname-mst-region] instance 1 vlan 2 to 10
[Sysname-mst-region] instance 2 vlan 20 to 30
[Sysname-mst-region] revision-level 1
[Sysname-mst-region] active region-configuration
```

Specifying the Root Bridge or a Secondary Root Bridge

MSTP can determine the root bridge of a spanning tree through MSTP calculation. Alternatively, you can specify the current device as the root bridge using the commands provided by the system.

Specifying the current device as the root bridge of a specific spanning tree

Follow these steps to specify the current device as the root bridge of a specific spanning tree:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Specify the current device as the root bridge of a specific spanning tree	stp [instance <i>instance-id</i>] root primary	Required By default, a device does not function as the root bridge.

Specifying the current device as a secondary root bridge of a specific spanning tree

Follow these steps to specify the current device as a secondary root bridge of a specific spanning tree:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Specify the current device as a secondary root bridge of a specific spanning tree	stp [instance <i>instance-id</i>] root secondary	Required By default, a device does not function as a secondary root bridge.



- Upon specifying the current device as the root bridge or a secondary root bridge, you cannot change the priority of the device.
- You can configure the current device as the root bridge or a secondary root bridge of an MST instance, which is specified by **instance** *instance-id* in the command. If you set *instance-id* to 0, the current device will be the root bridge or a secondary root bridge of the CIST.
- The current device has independent roles in different instances. It can act as the root bridge or a secondary root bridge of one instance while it can also act as the root bridge or a secondary root bridge of another instance. However, the same device cannot be the root bridge and a secondary root bridge in the same instance at the same time.
- There is one and only one root bridge in effect in a spanning tree instance. If two or more devices have been designated to be root bridges of the same spanning tree instance, MSTP will select the device with the lowest MAC address as the root bridge.
- You can specify multiple secondary root bridges for the same instance. Namely, you can specify secondary root bridges for the same instance on two or more than two devices.
- When the root bridge of an instance fails or is shut down, the secondary root bridge (if you have specified one) can take over the role of the instance. However, if you specify a new root bridge for the instance at this time, the secondary root bridge will not become the root bridge. If you have specified multiple secondary root bridges for an instance, when the root bridge fails,

MSTP will select the secondary root bridge with the lowest MAC address as the new root bridge.

- When specifying the root bridge or a secondary root bridge, you can specify the network diameter and hello time. However, these two options are effective only for MST instance 0, namely the CIST. If you include these two options in your command for any other instance, the configuration can succeed, but they will not actually work. For the description of network diameter and hello time, refer to “Configuring the Network Diameter of a Switched Network” on page 218 and “Configuring Timers of MSTP” on page 218.
- Alternatively, you can also specify the current device as the root bridge by setting the priority of the device to 0. For the device priority configuration, refer to “Configuring the Priority of the Current Device” on page 216.

Configuration example

Specify the current device as the root bridge of MST instance 1 and a secondary root bridge of MST instance 2.

```
<Sysname> system-view
[Sysname] stp instance 1 root primary
[Sysname] stp instance 2 root secondary
```

Configuring the Work Mode of MSTP Device

MSTP and RSTP can recognize each other’s protocol packets, so they are mutually compatible. However, STP is unable to recognize MSTP packets. For hybrid networking with legacy STP devices and full interoperability with RSTP-compliant devices, MSTP supports three work modes: STP-compatible mode, RSTP mode, and MSTP mode.

- In STP-compatible mode, all ports of the device send out STP BPDUs,
- In RSTP mode, all ports of the device send out RSTP BPDUs. If the device detects that it is connected with a legacy STP device, the port connecting with the legacy STP device will automatically migrate to STP-compatible mode.
- In MSTP mode, all ports of the device send out MSTP BPDUs. If the device detects that it is connected with a legacy STP device, the port connecting with the legacy STP device will automatically migrate to STP-compatible mode.

Configuration procedure

Follow these steps to configure the MSTP work mode:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the work mode of MSTP	stp mode { stp rstp mstp }	Optional MSTP mode by default

Configuration example

Configure MSTP to work in STP-compatible mode.

```
<Sysname> system-view
[Sysname] stp mode stp
```

Configuring the Priority of the Current Device

The priority of a device determines whether it can be elected as the root bridge of a spanning tree. A lower value indicates a higher priority. By setting the priority of

a device to a low value, you can specify the device as the root bridge of the spanning tree. An MSTP-compliant device can have different priorities in different MST instances.

Configuration procedure

Follow these steps to configure the priority of the current device:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the priority of the current device	stp [instance <i>instance-id</i>] priority <i>priority</i>	Optional 32768 by default



CAUTION:

- Upon specifying the current device as the root bridge or a secondary root bridge, you cannot change the priority of the device.
- During root bridge selection, if all devices in a spanning tree have the same priority, the one with the lowest MAC address will be selected as the root bridge of the spanning tree.

Configuration example

Set the device priority in MST instance 1 to 4096.

```
<Sysname> system-view
[Sysname] stp instance 1 priority 4096
```

Configuring the Maximum Hops of an MST Region

By setting the maximum hops of an MST region, you can restrict the region size. The maximum hops setting configured on the regional root bridge will be used as the maximum hops of the MST region.

The regional root bridge always sends a configuration BPDU with a hop count set to the maximum value. When a switch receives this configuration BPDU, it decrements the hop count by 1 and uses the new hop count as the remaining hop count in the BPDUs it propagates. When the hop count of a BPDU reaches 0, it is discarded by the device that received it. Thus, devices beyond the reach of the maximum hop are unable to take part in spanning tree calculation, and thereby the size of the MST region is confined.

When a device becomes the root bridge of the CIST or MSTI of an MST region, the maximum hop in the configuration BPDUs generated by this device defines the network diameter of the spanning tree to define how far the spanning tree can reach in this MST region. All the devices other than the root bridge in the MST region use the maximum hop value set for the root bridge.

Configuration procedure

Follow these steps to configure the maximum hops of the MST region:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the maximum hops of the MST region	stp max-hops <i>hops</i>	Optional 20 by default



A larger maximum hops setting means a larger size of the MST region. Only the maximum hops configured on the regional root bridge can restrict the size of the MST region.

Configuration example

Set the maximum hops of the MST region to 30.

```
<Sysname> system-view
[Sysname] stp max-hops 30
```

Configuring the Network Diameter of a Switched Network

Any two stations in a switched network are interconnected through specific paths, which are composed of a series of devices. Represented by the number of devices on a path, the network diameter is the path that comprises more devices than any other among these paths.

Configuration procedure

Follow these steps to configure the network diameter of the switched network:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the network diameter of the switched network	stp bridge-diameter <i>bridge-number</i>	Optional 7 by default



- *Network diameter is a parameter that indicates network size. A bigger network diameter represents a larger network size.*
- *Based on the network diameter you configured, MSTP automatically sets an optimal hello time, forward delay, and max age for the device.*
- *The configured network diameter is effective for the CIST only, and not for MSTIs.*

Configuration example

Set the network diameter of the switched network to 6.

```
<Sysname> system-view
[Sysname] stp bridge-diameter 6
```

Configuring Timers of MSTP

MSTP involves three timers: forward delay, hello time and max age. You can configure these three parameters for MSTP to calculate spanning trees.

Configuration procedure

Follow these steps to configure the timers of MSTP:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the forward delay timer	stp timer forward-delay <i>centi-seconds</i>	Optional 1,500 centiseconds (15 seconds) by default
Configure the hello time timer	stp timer hello <i>centi-seconds</i>	Optional 200 centiseconds (2 seconds) by default

To do...	Use the command...	Remarks
Configure the max age timer	stp timer max-age <i>centi-seconds</i>	Optional 2,000 centiseconds (20 seconds) by default

These three timers set on the root bridge of the CIST apply on all the devices on the entire switched network.



CAUTION:

- *The length of the forward delay time is related to the network diameter of the switched network. Typically, the larger the network diameter is, the longer the forward delay time should be. Note that if the forward delay setting is too small, temporary redundant paths may be introduced; if the forward delay setting is too big, it may take a long time for the network to resume connectivity. We recommend that you use the default setting.*
- *An appropriate hello time setting enables the device to timely detect link failures on the network without using excessive network resources. If the hello time is set too long, the device will take packet loss on a link for link failure and trigger a new spanning tree calculation process; if the hello time is set too short, the device will send repeated configuration BPDUs frequently, which adds to the device burden and causes waste of network resources. We recommend that you use the default setting.*
- *If the max age time setting is too small, the network devices will frequently launch spanning tree calculation and may take network congestion to a link failure; if the max age setting is too large, the network may fail to timely detect link failures and fail to timely launch spanning tree calculation, thus reducing the auto-sensing capability of the network. We recommend that you use the default setting.*

The setting of hello time, forward delay and max age must meet the following formulae; otherwise network instability will frequently occur.

- $2 \times (\text{forward delay} - 1 \text{ second}) \leq \text{max age}$
- $\text{Max age} \leq 2 \times (\text{hello time} + 1 \text{ second})$

We recommend that you specify the network diameter in the **stp root primary** command and let MSTP automatically calculate an optimal setting of these three timers.

Configuration example

Set the forward delay to 1,600 centiseconds, hello time to 300 centiseconds, and max age to 2,100 centiseconds.

```
<Sysname> system-view
[Sysname] stp timer forward-delay 1600
[Sysname] stp timer hello 300
[Sysname] stp timer max-age 2100
```

Configuring the Timeout Factor

After the network topology is stabilized, each non-root-bridge device forwards configuration BPDUs to the surrounding devices at the interval of hello time to check whether any link is faulty. Typically, if a device does not receive a BPDU from

the upstream device within nine times the hello time, it will assume that the upstream device has failed and start a new spanning tree calculation process.

In a very stable network, this kind of spanning tree calculation may occur because the upstream device is busy. In this case, you can avoid such unwanted spanning tree calculation by lengthening the timeout time.

Configuration procedure

Follow these steps to configure the timeout factor:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the timeout factor of the device	stp timer-factor <i>number</i>	Optional 3 by default



- *Timeout time = timeout factor × 3 × hello time.*
- *Typically, we recommend that you set the timeout factor to 5, or 6, or 7 for a stable network.*

Configuration example

Set the timeout factor to 6.

```
<Sysname> system-view
[Sysname] stp timer-factor 6
```

Configuring the Maximum Transmission Rate of Ports

The maximum transmission rate of a port refers to the maximum number of MSTP packets that the port can send within each hello time. The maximum transmission rate of an Ethernet port is related to the physical status of the port and the network structure.

Configuration procedure

Follow these steps to configure the maximum transmission rate of a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet interface view or port group view	Enter Ethernet interface view Enter port group view interface <i>interface-type</i> <i>interface-number</i> port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Required Use either command. Configurations made in Ethernet interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
Configure the maximum transmission rate of the port(s)	stp transmit-limit <i>packet-number</i>	Optional 10 by default



If the maximum transmission rate setting of a port is too big, the port will send a large number of MSTP packets within each hello time, thus using excessive network resources. We recommend that you use the default setting.

Configuration example

Set the maximum transmission rate of port GigabitEthernet 1/0/1 to 5.

```
<Sysname> system-view
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] stp transmit-limit 5
```

Configuring Ports as Edge Ports

If a port directly connects to a user terminal rather than another device or a shared LAN segment, this port is regarded as an edge port. When a network topology change occurs, an edge port will not cause a temporary loop. Therefore, if you specify a port as an edge port, this port can transition rapidly from the blocked state to the forwarding state without delay.

Configuration procedure

Follow these steps to specify a port or a group of ports as edge port(s):

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet interface view or port group view	Enter Ethernet interface view Enter port group view interface <i>interface-type interface-number</i> port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Required Use either command. Configurations made in Ethernet interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
Configure the port(s) as edge port(s)	stp edged-port enable	Required All Ethernet ports are non-edge ports by default.



- *With BPDU guard disabled, when a port set as an edge port receives a BPDU from another port, it will become a non-edge port again. In this case, you must reset the port before you can configure it to be an edge port again.*
- *If a port directly connects to a user terminal, configure it to be an edge port and enable BPDU guard for it. This enables the port to transition to the forwarding state while ensuring network security.*

Configuration example

Configure GigabitEthernet 1/0/1 to be an edge port.

```
<Sysname> system-view
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] stp edged-port enable
```

Configuring Whether Ports Connect to Point-to-Point Links

A point-to-point link is a link directly connecting with two devices. If the two ports across a point-to-point link are root ports or designated ports, the ports can rapidly transition to the forwarding state after a proposal-agreement handshake process.

Configuration procedure

Follow these steps to configure whether a port or a group of ports connect to point-to-point links:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet interface view or port group view	Enter Ethernet interface view Enter port group view interface <i>interface-type</i> <i>interface-number</i> port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Required Use either command. Configurations made in Ethernet interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
Configure whether the port(s) connect to point-to-point links	stp point-to-point { auto force-false force-true }	Optional The default setting is auto ; namely the device automatically detects whether an Ethernet port connects to a point-to-point link.



- *In the case of link aggregation, every port in the aggregation group can be configured to connect to a point-to-point link. If a port works in auto-negotiation mode and the negotiation result is full duplex, this port can be configured as connecting to a point-to-point link.*
- *If a port is configured as connecting to a point-to-point link, the setting takes effect for the port in all MST instances. If the physical link to which the port connects is not a point-to-point link and you force it to be a point-to-point link by configuration, the configuration may incur a temporary loop.*

Configuration example

Configure port GigabitEthernet 1/0/1 as connecting to a point-to-point link.

```
<Sysname> system-view
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] stp point-to-point force-true
```

Configuring the Mode a Port Uses to Recognize/Send MSTP Packets

A port can send/recognize MSTP packets of two formats:

- 802.1s-compliant standard format, and
- Compatible format

By default, the packet format recognition mode of a port is **auto**, namely the port automatically distinguishes the two MSTP packet formats, and determines the format of packets it will send based on the recognized format. You can configure the MSTP packet format to be used by a port. After the configuration, when working in MSTP mode, the port sends and receives only MSTP packets of the format you have configured to communicate with devices that send the same format of packets.

Configuration procedure

Follow these steps to configure the MSTP packet format to be supported by a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet interface view or port group view	interface <i>interface-type</i> <i>interface-number</i>	Required Use either command. Configurations made in Ethernet interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
Configure the mode the port uses to recognize/send MSTP packets	port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Optional auto by default
	stp compliance { auto dot1s legacy }	Optional auto by default



- In MSTP mode, if a port is configured to recognize/send MSTP packets in a mode other than **auto**, and if it receives a packet in the format different from the specified type, that port will become a designated port and remain in the discarding state to prevent the occurrence of a loop.
- If a port receives MSTP packets of different formats frequently, this means that the MSTP packet formation configuration contains error. In this case, if the port is working in MSTP mode, it will be disabled for protection. Those ports closed thereby can be restored only by the network administrators.

Configuration example

Configure GigabitEthernet 1/0/1 to receive and send standard-format MSTP packets.

```
<Sysname> system-view
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] stp compliance dot1s
```

Enabling the Output of Port State Transition Information

In a large-scale, MSTP-enabled network, there are a large number of MSTP instances, so ports may frequently transition from one state to another. In this situation, you can enable the device to output the port state transition information of all SPT instances or the specified SPT instance so as to monitor the port states in real time.

Follow these steps to enable output of port state transition information:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable output of port state transition information of all instances or a particular instance	stp port-log { all instance <i>instance-id</i> }	Optional Enabled by default

Enabling the MSTP Feature

Configuration procedure

Follow these steps to enable the MSTP feature:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the MSTP feature for the device	stp enable	Required Disabled by default
Enter Ethernet interface view or port group view	Enter Ethernet interface view Enter port group view	Required Use either command. Configurations made in Ethernet interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
Enable the MSTP feature on the port(s)	stp enable	Optional MSTP is disabled on ports by default and automatically enabled on all ports after it is enabled globally on the device.



- You must enable MSTP for the device before any other MSTP-related configuration can take effect.
- To control MSTP flexibly, you can use the **stp disable** or **undo stp** command to disable the MSTP feature for certain ports so that they will not take part in spanning tree calculation and thus to save the device's CPU resources.

Configuration example

Enable MSTP for the device and disable MSTP on port GigabitEthernet1/0/1.

```
<Sysname> system-view
[Sysname] stp enable
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] stp disable
```

Configuring Leaf Nodes

Configuring an MST Region

Refer to "Configuring an MST Region" on page 213 in the section about root bridge configuration.

Configuring the Work Mode of MSTP

Refer to "Configuring the Work Mode of MSTP Device" on page 216 in the section about root bridge configuration.

Configuring the Timeout Factor

Refer to "Configuring Timers of MSTP" on page 218 in the section about root bridge configuration.

Configuring the Maximum Transmission Rate of Ports

Refer to “Configuring the Maximum Transmission Rate of Ports” on page 220 in the section about root bridge configuration.

Configuring Ports as Edge Ports

Refer to “Configuring Ports as Edge Ports” on page 221 in the section about root bridge configuration.

Configuring Path Costs of Ports

Path cost is a parameter related to the rate of port-connected links. On an MSTP-compliant device, ports can have different priorities in different MST instances. Setting an appropriate path cost allows VLAN traffic flows to be forwarded along different physical links, thus to enable per-VLAN load balancing.

The device can automatically calculate the default path cost; alternatively, you can also configure the path cost for ports.

Specifying a standard that the device uses when calculating the default path cost

You can specify a standard for the device to use in automatic calculation for the default path cost. The device supports the following standards:

- **dot1d-1998**: The device calculates the default path cost for ports based on IEEE 802.1d-1998.
- **dot1t**: The device calculates the default path cost for ports based on IEEE 802.1t.
- **legacy**: The device calculates the default path cost for ports based on a private standard.

Follow these steps to specify a standard for the device to use when calculating the default path cost:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Specify a standard for the device to use when calculating the default path cost of the link connected with the device	stp pathcost-standard { dot1d-1998 dot1t legacy }	Optional legacy by default

Table 43 Link speed vs. path cost

Link speed	Duplex state	802.1d-1998	802.1t	Private standard
0	-	65535	200,000,000	200,000
10 Mbps	Single Port	100	2,000,000	2,000
	Aggregated Link 2 Ports	100	1,000,000	1,800
	Aggregated Link 3 Ports	100	666,666	1,600
	Aggregated Link 4 Ports	100	500,000	1,400
100 Mbps	Single Port	19	200,000	200
	Aggregated Link 2 Ports	19	100,000	180
	Aggregated Link 3 Ports	19	66,666	160
	Aggregated Link 4 Ports	19	50,000	140

Table 43 Link speed vs. path cost

Link speed	Duplex state	802.1d-1998	802.1t	Private standard
1000 Mbps	Single Port	4	20,000	20
	Aggregated Link 2 Ports	4	10,000	18
	Aggregated Link 3 Ports	4	6,666	16
	Aggregated Link 4 Ports	4	5,000	14
10 Gbps	Single Port	2	2,000	2
	Aggregated Link 2 Ports	2	1,000	1
	Aggregated Link 3 Ports	2	666	1
	Aggregated Link 4 Ports	2	500	1



In the calculation of the path cost value of an aggregated link, 802.1d-1998 does not take into account the number of ports in the aggregated link. Whereas, 802.1t takes the number of ports in the aggregated link into account. The calculation formula is: $\text{Path Cost} = 200,000,000 / \text{link speed (in 100 kbps)}$, where link speed is the sum of the link speed values of the non-blocked ports in the aggregated link.

Configuring Path Costs of Ports

Follow these steps to configure the path cost of ports:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet interface view or port group view	Enter Ethernet interface view interface <i>interface-type</i> <i>interface-number</i> Enter port group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Required Use either command. Configurations made in Ethernet interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
Configure the path cost of the port(s)	stp [instance <i>instance-id</i>] cost <i>cost</i>	Required By default, MSTP automatically calculates the path cost of each port.



CAUTION:

- If you change the standard that the device uses in calculating the default path cost, the port path cost value set through the **stp cost** command will be out of effect.
- When the path cost of a port is changed, MSTP will re-calculate the role of the port and initiate a state transition. If you use 0 as *instance-id*, you are setting the path cost of the CIST.

Configuring Port Priority

The priority of a port is an import basis that determines whether the port can be elected as the root port of device. If all other conditions are the same, the port with the highest priority will be elected as the root port.

On an MSTP-compliant device, a port can have different priorities in different MST instances, and the same port can play different roles in different MST instances, so that data of different VLANs can be propagated along different physical paths, thus implementing per-VLAN load balancing. You can set port priority values based on the actual networking requirements.

Configuration procedure

Follow these steps to configure the priority of a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet interface view or port group view	Enter Ethernet interface view interface <i>interface-type interface-number</i> Enter port group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Required Use either command. Configurations made in Ethernet interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
Configure the port priority	stp [instance <i>instance-id</i>] port priority <i>priority</i>	Optional 128 for all Ethernet ports by default.



- *When the priority of a port is changed, MSTP will re-calculate the role of the port and initiate a state transition.*
- *Generally, a lower configured value priority indicates a higher priority of the port. If you configure the same priority value for all the Ethernet ports on a device, the specific priority of a port depends on the index number of that port. Changing the priority of an Ethernet port triggers a new spanning tree calculation process.*

Configuration example

Set the priority of port GigabitEthernet 1/0/1 to 16 in MST instance 1.

```
<Sysname> system-view
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] stp instance 1 port priority 16
```

Configuring Whether Ports Connect to Point-to-Point Links

Refer to “Configuring Whether Ports Connect to Point-to-Point Links” on page 221 in the section about root bridge configuration.

Configuring the Mode a Port Uses to Recognize/Send MSTP Packets

Refer to “Configuring the Mode a Port Uses to Recognize/Send MSTP Packets” on page 222 in the section about root bridge configuration.

Enabling Output of Port State Transition Information

Refer to “Enabling the Output of Port State Transition Information” on page 223 in the section about root bridge configuration.

Enabling the MSTP Feature

Refer to “Enabling the MSTP Feature” on page 224 in the section about root bridge configuration.

Performing mCheck

Ports on an MSTP-compliant device have three working modes: STP compatible mode, RSTP mode, and MSTP mode.

In a switched network, if a port on the device running MSTP (or RSTP) connects to a device running STP, this port will automatically migrate to the STP-compatible mode. However, if the device running STP is removed, this will not be able to migrate automatically to the MSTP (or RSTP) mode, but will remain working in the STP-compatible mode. In this case, you can perform an mCheck operation to force the port to migrate to the MSTP (or RSTP) mode.

You can perform mCheck on a port through two approaches, which lead to the same result.

Configuration Prerequisites

MSTP has been correctly configured on the device.

Configuration Procedure**Performing mCheckglobally**

Follow these steps to perform global mCheck:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Perform mCheck	stp mcheck	Required

Performing mCheck in Ethernet interface view

Follow these steps to perform mCheck in Ethernet interface view:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet interface view	interface <i>interface-type interface-number</i>	-
Perform mCheck	stp mcheck	Required



CAUTION: The **stp mcheck** command is meaningful only when the device works in the MSTP (or RSTP) mode, not in the STP-compatible mode.

Configuration Example

Perform mCheck on port GigabitEthernet 1/0/1.

1 Method 1: Perform mCheck globally.

```
<Sysname> system-view
[Sysname] stp mcheck
```

2 Method 2: Perform mCheck in Ethernet interface view.

```
<Sysname> system-view
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] stp mcheck
```


Configuring Digest Snooping

As defined in IEEE 802.1s, interconnected devices are in the same region only when the region-related configuration (domain name, revision level, VLAN-to-instance mappings) on them is identical. An MSTP-enabled device identifies devices in the same MST region by checking the configuration ID in BPDU packets. The configuration ID includes the region name, revision level, configuration digest that is in 16-byte length and is the result calculated via the HMAC-MD5 algorithm based on VLAN-to-instance mappings.

Since MSTP implementations differ with vendors, the configuration digest calculated using private key is different; hence different vendors' devices in the same MST region can not communicate with each other.

Enabling the Digest Snooping feature on the associated port can make a device communicate with another vendor's device in the same MST region.

Configuration Prerequisites

Associated devices of different vendors are interconnected and run MSTP.

Configuration Procedure

Follow these steps to configure Digest Snooping:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet interface or port group view	Enter Ethernet interface view Enter port group view	Required Use either command. Configurations made in Ethernet interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
Enable digest snooping on the interface or port group	stp config-digest-snooping	Required Not enabled by default
Return to system view	quit	-
Enable global digest snooping	stp config-digest-snooping	Required Not enabled by default



CAUTION:

- You can only enable the Digest Snooping feature on the device connected to another vendor's device that uses a private key to calculate the configuration digest.
- With the Digest Snooping feature enabled, comparison of configuration digest is not needed for in-the-same-region check, so the VLAN-to-instance mappings must be the same on associated ports.
- With global Digest Snooping enabled, modification of VLAN-to-instance mappings and removing of the current region configuration using the **undo stp region-configuration** command are not allowed. You can only modify the region name and revision level.

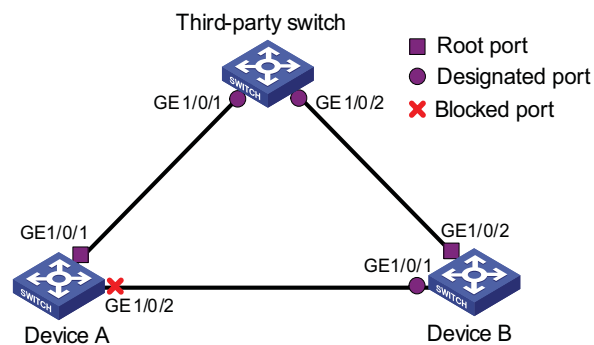
- You need to enable this feature both globally and on associated ports to make it take effect. It is recommended to enable the feature on all associated ports first and then globally, making all configured ports take effect, and disable the feature globally to disable it on all associated ports.
- It is not recommended to enable Digest Snooping on the MST region edge port to avoid loops.
- It is recommended to enable Digest Snooping first and then MSTP. Do not enable Digest Snooping when the network works well to avoid traffic interruption.

Configuration Example Network requirements

- Device A and Device B connect to a third-party's router and all the routers are in the same region.
- Enable Digest Snooping on Device A and Device B so that the three routers can communicate with one another.

Network diagram

Figure 58 Digest Snooping configuration



Configuration procedure

1 Enable Digest Snooping on Device A

Enable Digest Snooping on GigabitEthernet1/0/1.

```
<DeviceA> system-view
[DeviceA] interface GigabitEthernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] stp config-digest-snooping
```

Enable global Digest Snooping.

```
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] stp config-digest-snooping
```

2 Enable Digest Snooping on Device B (the same as above, omitted)

Configuring No Agreement Check

Two types of messages are used for rapid state transition on designated RSTP and MSTP ports:

- Proposal: sent by designated ports to request rapid transition
- Agreement: used to acknowledge rapid transition requests

Both RSTP and MSTP switches can perform rapid transition operation on a designated port only when the port receives an agreement packet from the downstream switch. The differences between RSTP and MSTP switches are:

- For MSTP, the downstream device's root port sends an agreement packet only after it receives an agreement packet from the upstream device.
- For RSTP, the downstream device sends an agreement packet regardless of whether an agreement packet from the upstream device is received.

Figure 59 and Figure 60 show the rapid state transition mechanism on MSTP and RSTP designated ports.

Figure 59 Rapid state transition of a designated port in MSTP

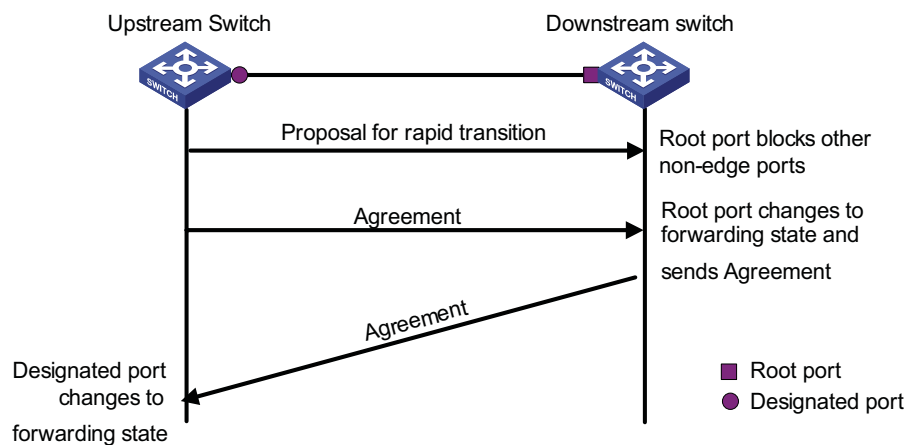
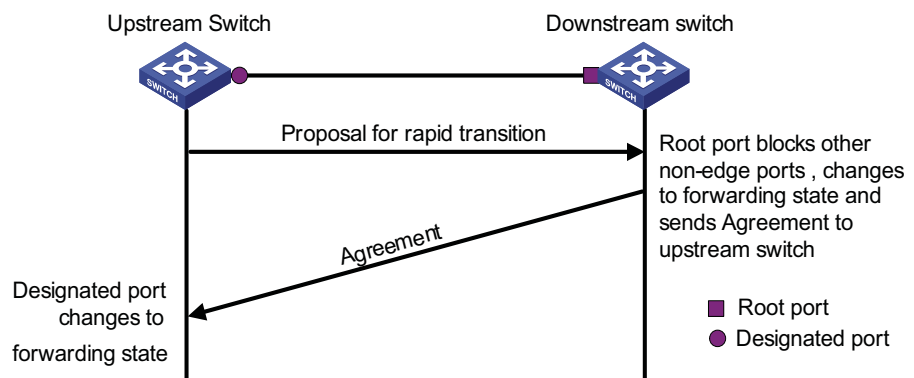


Figure 60 Rapid state transition of a designated port in RSTP



If the upstream device comes from another vendor, the rapid state transition implementation may be limited. For example, when the upstream device adopts RSTP, the downstream device adopts MSTP and does not support RSTP mode, the root port on the downstream device receives no agreement packet from the upstream device and thus sends no agreement packets to the upstream device. As a result, the designated port of the upstream switch fails to transit rapidly and can only change to the forwarding state after a period twice the Forward Delay.

In this case, you can enable the No Agreement Check feature on the downstream device's port to perform rapid state transition.

- Prerequisites**
- A device is the upstream one that is connected to another vendor's MSTP supported device via a point-to-point link.
 - Configure the same region name, revision level and VLAN-to-instance mappings on the two devices, making them in the same region.

Configuration Procedure Follow these steps to configure No Agreement Check:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet interface or port group view	Enter Ethernet interface view interface <i>interface-type</i> <i>interface-number</i> Enter port group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Required Use either command. Configurations made in Ethernet interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
Enable No Agreement Check	stp no-agreement-check	Required Not enabled by default



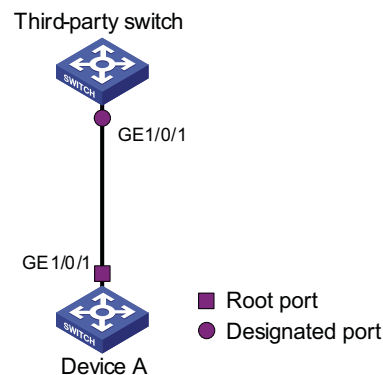
The No Agreement Check feature can only take effect on the root port or Alternate port after enabled.

Configuration Example Network requirements

- Device A connects to a third-party's device that has different MSTP implementation. Both switches are in the same region.
- Another vendor's device is the regional root bridge, and Device A is the downstream device.

Network diagram

Figure 61 No Agreement Check configuration



Configuration procedure

Enable No Agreement Check on GigabitEthernet1/0/1 of Device A.

```
<DeviceA> system-view
[DeviceA] interface GigabitEthernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] stp no-agreement-check
```

Configuring Protection Functions

An MSTP-compliant device supports the following protection functions:

- BPDU guard
- Root guard
- Loop guard
- TC-BPDU attack guard



- *The the Switch 4800G support the BPDU guard, root guard and loop guard functions.*
- *Among loop guard, root guard and edge port setting, only one function can take effect on the same port at the same time.*

Configuration prerequisites

MSTP has been correctly configured on the device.

Enabling BPDU Guard

For access layer devices, the access ports generally connect directly with user terminals (such as PCs) or file servers. In this case, the access ports are configured as edge ports to allow rapid transition of these ports. When these ports receive configuration BPDUs, the system will automatically set these ports as non-edge ports and start a new spanning tree calculation process. This will cause a change of network topology. Under normal conditions, these ports should not receive configuration BPDUs. However, if someone forges configuration BPDUs maliciously to attack the devices, network instability will occur.

MSTP provides the BPDU guard function to protect the system against such attacks. With the BPDU guard function enabled on the devices, when edge ports receive configuration BPDUs, MSTP will close these ports and notify the NMS that these ports have been closed by MSTP. Those ports closed thereby can be restored only by the network administrators.



It is recommended that you enable the BPDU guard on your device.

Follow these steps to enable BPDU guard:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the BPDU guard function on the device	stp bpdu-protection	Required Disabled by default

Enabling Root Guard

The root bridge and secondary root bridge of a panning tree should be located in the same MST region. Especially for the CIST, the root bridge and secondary root bridge are generally put in a high-bandwidth core region during network design. However, due to possible configuration errors or malicious attacks in the network, the legal root bridge may receive a configuration BPDU with a higher priority. In this case, the current, legal root bridge will be superseded by another device, causing undesired change of the network topology. As a result of this kind of illegal topology change, the traffic that should go over high-speed links is drawn to low-speed links, resulting in network congestion.

To prevent this situation from happening, MSTP provides the root guard function to protect the root bridge. If the root guard function is enabled on a port, this port will keep playing the role of designated port on all MST instances. Once this port receives a configuration BPDU with a higher priority from an MST instance, it immediately sets that instance port to the listening state, without forwarding the packet (this is equivalent to disconnecting the link connected with this port). If the port receives no BPDUs with a higher priority within twice the forwarding delay, the port will revert to its original state.



It is recommended that you enable the root guard feature on your device.

Follow these steps to enable root guard:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet interface view or port group view	Enter Ethernet interface view interface <i>interface-type</i> <i>interface-number</i> Enter port group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Required Use either command. Configurations made in Ethernet interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
Enable the root guard function on the port(s)	stp root-protection	Required Disabled by default

Enabling Loop Guard

By keeping receiving BPDUs from the upstream device, a device can maintain the state of the root port and other blocked ports. However, due to link congestion or unidirectional link failures, these ports may fail to receive BPDUs from the upstream device. In this case, the downstream device will reselect the port roles: those ports failed to receive upstream BPDUs will become designated ports and the blocked ports will transition to the forwarding state, resulting in loops in the switched network. The loop guard function can suppress the occurrence of such loops.

If a loop guard-enabled port fails to receive BPDUs from the upstream device, and if the port took part in STP calculation, all the instances on the port, no matter what roles they play, will be set to, and stay in, the Discarding state.



It is recommended that you enable the loop guard feature on your device.

Follow these steps to enable loop guard:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks	
Enter Ethernet interface view or port group view	Enter Ethernet interface view Enter port group view	interface <i>interface-type interface-number</i> port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Required Use either command. Configurations made in Ethernet interface view will take effect on the current port only; configurations made in port group view will take effect on all ports in the port group.
Enable the loop guard function for the port(s)	stp loop-protection	Required Disabled by default	

Enabling TC-BPDU Attack Guard

When receiving a TC-BPDU (a PDU used as notification of topology change), the device will delete the corresponding forwarding address entry. If someone forges TC-BPDUs to attack the device, the device will receive a larger number of TC-BPDUs within a short time, and frequent deletion operations bring a big burden to the device and hazard network stability.

With the TC-BPDU guard function enabled, the device limits the maximum number of times of immediately deleting forwarding address entries within 10 seconds after it receives TC-BPDUs to the value set with the **stp tc-protection threshold** command (assume the value is X). At the same time, the system monitors whether the number of TC-BPDUs received within that period of time is larger than X. If so, the device will perform another deletion operation after that period of time elapses. This prevents frequent deletion of forwarding address entries.

Follow these steps to enable TC-BPDU attack guard:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the TC-BPDU attack guard function	stp tc-protection enable	Optional Enabled by default
Configure the maximum number of times the device deletes forwarding address entries within a certain period of time immediately after it receives TC-BPDUs	stp tc-protection threshold <i>number</i>	Optional 6 by default



We recommend that you keep this feature enabled.

Displaying and Maintaining MSTP

To do...	Use the command...	Remarks
View the information about abnormally blocked ports	display stp abnormal-port	Available in any view
View the information about ports blocked by STP protection actions	display stp down-port	Available in any view

To do...	Use the command...	Remarks
View the information of port role calculation history for the specified MSTP instance or all MSTP instances	display stp [instance <i>instance-id</i>] history	Available in any view
View the statistics of TC/TCN BPDUs sent and received by all ports in the specified MSTP instance or all MSTP instances	display stp [instance <i>instance-id</i>] tc	Available in any view
View the status information and statistics information of MSTP	display stp [instance <i>instance-id</i>] [interface <i>interface-list</i>] [brief]	Available in any view
View the information about MSTP region configuration in effect	display stp region-configuration	Available in any view
View root bridge information of all MSTP instances	display stp root	Available in any view
Clear the statistics information of MSTP	reset stp [interface <i>interface-list</i>]	Available in user view

MSTP Configuration Example

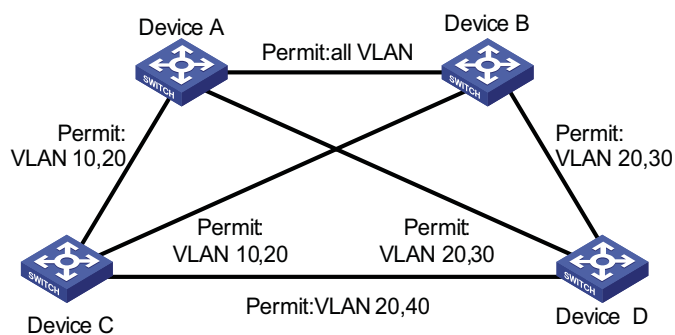
Network requirements

Configure MSTP so that packets of different VLANs are forwarded along different spanning trees. The specific configuration requirements are as follows:

- All devices on the network are in the same MST region.
- Packets of VLAN 10 are forwarded along MST region 1, those of VLAN 30 are forwarded along MST instance 3, those of VLAN 40 are forwarded along MST instance 4, and those of VLAN 20 are forwarded along MST instance 0.
- Device A and Device B are convergence layer devices, while Device C and Device D are access layer devices. VLAN 10 and VLAN 30 are terminated on the convergence layer devices, and VLAN 40 is terminated on the access layer devices, so the root bridges of MST instance 1 and MST instance 3 are Device A and Device B respectively, while the root bridge of MST instance 4 is Device C.

Network diagram

Figure 62 Network diagram for MSTP configuration



“Permit:” beside each link in the figure is followed by the VLANs the packets of which are permitted to pass this link.

Configuration procedure

1 Configuration on Device A

Enter MST region view.

```
<DeviceA> system-view
[DeviceA] stp region-configuration
```

Configure the region name, VLAN-to-instance mappings and revision level of the MST region.

```
[DeviceA-mst-region] region-name example
[DeviceA-mst-region] instance 1 vlan 10
[DeviceA-mst-region] instance 3 vlan 30
[DeviceA-mst-region] instance 4 vlan 40
[DeviceA-mst-region] revision-level 0
```

Activate MST region configuration manually.

```
[DeviceA-mst-region] active region-configuration
[DeviceA-mst-region] quit
```

Define Device A as the root bridge of MST instance 1.

```
[DeviceA] stp instance 1 root primary
```

View the MST region configuration information that has taken effect.

```
[DeviceA] display stp region-configuration
Oper configuration
Format selector      :0
Region name         :example
Revision level      :0

Instance   Vlans Mapped
-----
0          1 to 9, 11 to 29, 31 to 39, 41 to 4094
1          10
3          30
4          40
```

2 Configuration on Device B

Enter MST region view.

```
<DeviceB> system-view
[DeviceB] stp region-configuration
```

Configure the region name, VLAN-to-instance mappings and revision level of the MST region.

```
[DeviceB-mst-region] region-name example
[DeviceB-mst-region] instance 1 vlan 10
[DeviceB-mst-region] instance 3 vlan 30
[DeviceB-mst-region] instance 4 vlan 40
[DeviceB-mst-region] revision-level 0
```

Activate MST region configuration manually.

```
[DeviceB-mst-region] active region-configuration
[DeviceB-mst-region] quit
```

Define Device B as the root bridge of MST instance 3.

```
[DeviceB] stp instance 3 root primary
```

View the MST region configuration information that has taken effect.

```
[DeviceB] display stp region-configuration
Oper configuration
  Format selector      :0
  Region name         :example
  Revision level      :0

Instance  Vlans Mapped
   0       1 to 9, 11 to 29, 31 to 39, 41 to 4094
   1         10
   3         30
   4         40
```

3 Configuration on Device C

Enter MST region view.

```
<DeviceC> system-view
[DeviceC] stp region-configuration
[DeviceC-mst-region] region-name example
```

Configure the region name, VLAN-to-instance mappings and revision level of the MST region.

```
[DeviceC-mst-region] instance 1 vlan 10
[DeviceC-mst-region] instance 3 vlan 30
[DeviceC-mst-region] instance 4 vlan 40
[DeviceC-mst-region] revision-level 0
```

Activate MST region configuration manually.

```
[DeviceC-mst-region] active region-configuration
[DeviceC-mst-region] quit
```

Define Device C as the root bridge of MST instance 4.

```
[DeviceC] stp instance 4 root primary
```

View the MST region configuration information that has taken effect.

```
[DeviceC] display stp region-configuration
Oper configuration
  Format selector      :0
  Region name         :example
  Revision level      :0

Instance  Vlans Mapped
   0       1 to 9, 11 to 29, 31 to 39, 41 to 4094
   1         10
   3         30
   4         40
```

4 Configuration on Device D

Enter MST region view.

```
<DeviceD> system-view
[DeviceD] stp region-configuration
[DeviceD-mst-region] region-name example
```

Configure the region name, VLAN-to-instance mappings and revision level of the MST region.

```
[DeviceD-mst-region] instance 1 vlan 10
[DeviceD-mst-region] instance 3 vlan 30
[DeviceD-mst-region] instance 4 vlan 40
[DeviceD-mst-region] revision-level 0
```

Activate MST region configuration manually.

```
[DeviceD-mst-region] active region-configuration
[DeviceD-mst-region] quit
```

View the MST region configuration information that has taken effect.

```
[DeviceD] display stp region-configuration
Oper configuration
  Format selector      :0
  Region name         :example
  Revision level      :0

Instance   Vlans Mapped
  0         1 to 9, 11 to 29, 31 to 39, 41 to 4094
  1         10
  3         30
  4         40
```


24

IP ROUTING OVERVIEW

Go to these sections for information you are interested in:

- "IP Routing and Routing Table" on page 241
- "Routing Protocol Overview" on page 243
- "Displaying and Maintaining a Routing Table" on page 246



The term "router" in this document refers to a Layer 3 switch running routing protocols.

IP Routing and Routing Table

Routing Routing in the Internet is achieved through routers. Upon receiving a packet, a router finds an optimal route based on the destination address and forwards the packet to the next router in the path until the packet reaches the last router, which forwards the packet to the intended destination host.

Routing Through a Routing Table

Routing table

Routing tables play a key role in routing. Each router maintains a routing table, and each entry in the table specifies which physical interface a packet destined for a certain destination should go out to reach the next hop (the next router) or the directly connected destination.

Routes in a routing table can be divided into three categories by origin:

- Direct routes: Routes discovered by data link protocols, also known as interface routes.
- Static routes: Routes that are manually configured.
- Dynamic routes: Routes that are discovered dynamically by routing protocols.

Contents of a routing table

A routing table includes the following key items:

- Destination address: Destination IP address or destination network.
- Network mask: Specifies, in company with the destination address, the address of the destination network. A logical AND operation between the destination address and the network mask yields the address of the destination network. For example, if the destination address is 129.102.8.10 and the mask 255.255.0.0, the address of the destination network is 129.102.0.0. A network mask is made of a certain number of consecutive 1s. It can be expressed in dotted decimal format or by the number of the 1s.

- Outbound interface: Specifies the interface through which the IP packets are to be forwarded.
- IP address of the next hop: Specifies the address of the next router on the path. If only the outbound interface is configured, its address will be the IP address of the next hop.
- Priority for the route. Routes to the same destination but having different nexthops may have different priorities and be found by various routing protocols or manually configured. The optimal route is the one with the highest priority (with the smallest metric).

Routes can be divided into two categories by destination:

- Subnet routes: The destination is a subnet.
- Host routes: The destination is a host.

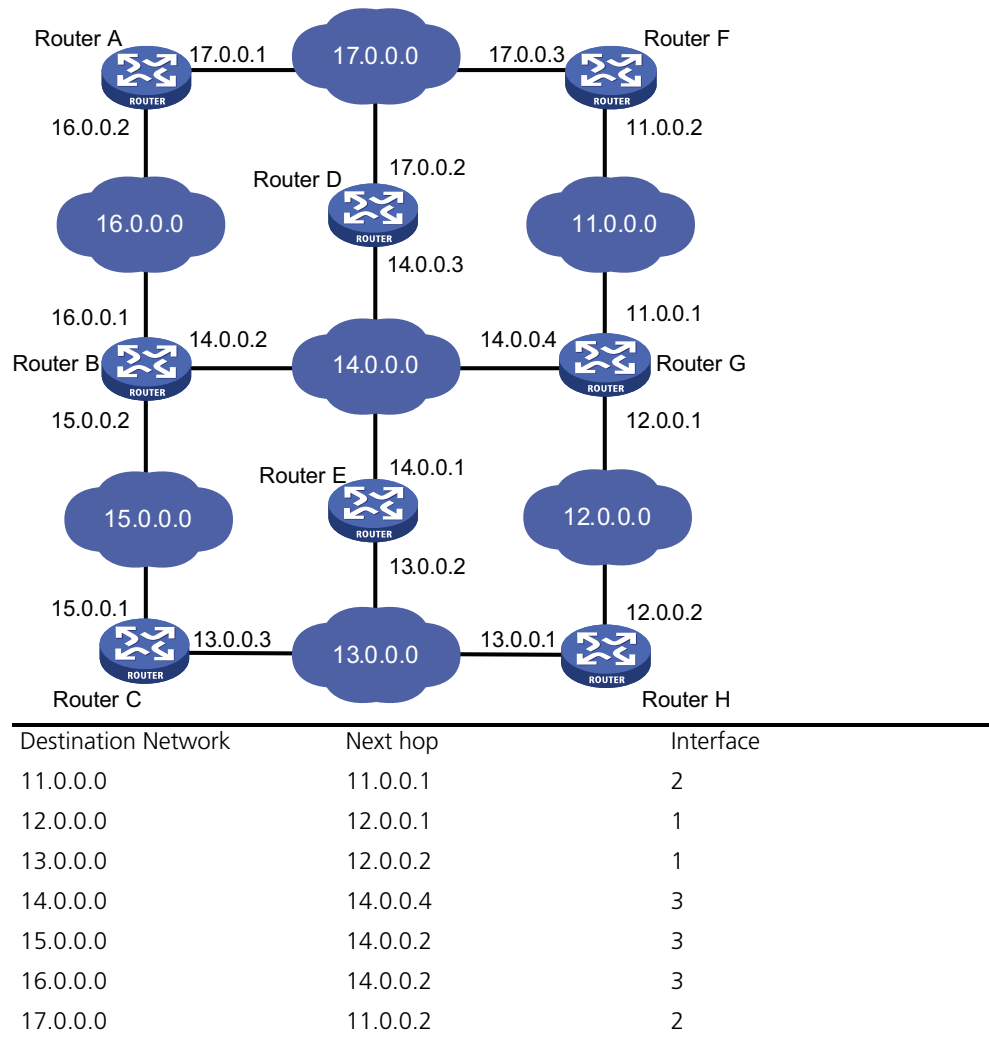
Based on whether the destination is directly connected to a given router, routes can be divided into:

- Direct routes: The destination is directly connected to the router.
- Indirect routes: The destination is not directly connected to the router.

To prevent the routing table from getting too large, you can configure a default route. All packets without matching entry in the routing table will be forwarded through the default route.

In Figure 63, the IP address on each cloud represents the address of the network. Router G resides in three networks and therefore has three IP addresses for its three physical interfaces. Its routing table is shown on the right of the network topology.

Figure 63 A sample routing table



Routing Protocol Overview

Static Routing and Dynamic Routing

Static routing is easy to configure and requires less system resources. It works well in small, stable networks with simple topologies. Its major drawback is that you must perform routing configuration again whenever the network topology changes; it cannot adjust to network changes by itself.

Dynamic routing is based on dynamic routing protocols, which can detect network topology changes and recalculate the routes accordingly. Therefore, dynamic routing is suitable for large networks. Its disadvantages are that it is complicated to configure, and that it not only imposes higher requirements on the system, but also eats away a certain amount of network resources.

Classification of Dynamic Routing Protocols

Dynamic routing protocols can be classified based on the following standards:

Operational scope

- Interior gateway protocols (IGPs): Work within an autonomous system, including RIP, OSPF, and IS-IS.
- Exterior gateway protocols (EGPs): Work between autonomous systems. The most popular one is BGP.



An autonomous system refers to a group of routers that share the same routing policy and work under the same administration.

Routing algorithm

- Distance-vector protocols: RIP and BGP. BGP is also considered a path-vector protocol.
- Link-state protocols: OSPF and IS-IS.

The main differences between the above two types of routing algorithms lie in the way routes are discovered and calculated.

Type of the destination address

- Unicast routing protocols: RIP, OSPF, BGP, and IS-IS.
- Multicast routing protocols: PIM-SM and PIM-DM.

This chapter focuses on unicast routing protocols. For information on multicast routing protocols, refer to the “Multicast Routing and Forwarding Overview” on page 701.

Version of IP protocol

IPv4 routing protocols: RIP, OSPFv2, BGP4 and IS-IS.

IPv6 routing protocols: RIPng, OSPFv3, IPv6 BGP, and IPv6 IS-IS.

Routing Protocols and Routing Priority

Different routing protocols may find different routes to the same destination. However, not all of those routes are optimal. In fact, at a particular moment, only one protocol can uniquely determine the current optimal routing to the destination. For the purpose of route selection, each routing protocol (including static routes) is assigned a priority. The route found by the routing protocol with the highest priority is preferred.

The following table lists some routing protocols and the default priorities for routes found by them:

Routing approach	Priority
DIRECT	0
OSPF	10
IS-IS	15
STATIC	60
RIP	100
OSPF ASE	150
OSPF NSSA	150
IBGP	255

Routing approach	Priority
EBGP	255
UNKNOWN	256



- *The smaller the priority value, the higher the priority.*
- *The priority for a direct route is always 0, which you cannot change. Any other type of routes can have their priorities manually configured.*
- *Each static route can be configured with a different priority.*
- *IPv4 and IPv6 routes have their own respective routing tables.*

Load Balancing and Route Backup

Load balancing

In multi-route mode, a routing protocol can be configured with multiple equal-cost routes to the same destination. These routes have the same priority and will all be used to accomplish load balancing if there is no route with a higher priority available.

A given routing protocol may find several routes with the same metric to the same destination, and if this protocol has the highest priority among all the active protocols, these routes will be considered valid routes for load balancing.

In current implementations, routing protocols supporting load balancing are static routing, RIP, OSPF, BGP and IS-IS.

Route backup

Route backup can help improve network reliability. With route backup, you can configure multiple routes to the same destination, expecting the one with the highest priority to be the main route and all the rest backup routes.

Under normal circumstances, packets are forwarded through the main route. When the main route goes down, the route with the highest priority among the backup routes is selected to forward packets. When the main route recovers, the route selection process is performed again and the main route is selected again to forward packets.

Route Recursion

The nexthops of some BGP routes (except EBGP routes) and static routes configured with nexthops may not be directly connected. To forward the packets, the outgoing interface to reach the nexthop must be available. Route recursion is used to find the outgoing interface based on the nexthop information of the route. Link-state routing protocols, such as OSPF and IS-IS, do not need route recursion because they obtain nexthop information through route calculation.

Sharing of Routing Information

As different routing protocols use different routing algorithms to calculate routes, they may find different routes. In a large network with multiple routing protocols, it is required for routing protocols to share their routing information. Each routing protocol has its own route redistribution mechanism. For detailed information, refer to the description about route redistribution in each routing protocol.

Displaying and Maintaining a Routing Table

To do...	Use the command...	Remarks
Display brief information about the active routes in the routing table	display ip routing-table [verbose] [{ begin exclude include } <i>regular-expression</i>]	Available in any view
Display information about routes to the specified destination	display ip routing-table <i>ip-address</i> [<i>mask-length</i> <i>mask</i>] [longer-match] [verbose]	
Display information about routes with destination addresses in the specified range	display ip routing-table <i>ip-address1</i> { <i>mask-length</i> <i>mask</i> } <i>ip-address2</i> { <i>mask-length</i> <i>mask</i> } [verbose]	
Display information about routes permitted by an IPv4 basic ACL	display ip routing-table acl <i>acl-number</i> [verbose]	
Display routing information permitted by an IPv4 prefix list	display ip routing-table ip-prefix <i>ip-prefix-name</i> [verbose]	Available in any view
Display routes of a routing protocol	display ip routing-table protocol <i>protocol</i> [inactive verbose]	
Display statistics about the network routing table	display ip routing-table statistics	
Clear statistics for the routing table	reset ip routing-table statistics protocol { all <i>protocol</i> }	Available in user view
Display the information of recursive routes	display ip relay-route	Available in any view
Display IPv6 recursive route information	display ipv6 relay-route	
Display brief IPv6 routing table information	display ipv6 routing-table	
Display verbose IPv6 routing table information	display ipv6 routing-table verbose	
Display routing information for a specified destination IPv6 address	display ipv6 routing-table <i>ipv6-address</i> <i>prefix-length</i> [longer-match] [verbose]	
Display routing information permitted by an IPv6 ACL	display ipv6 routing-table acl <i>acl6-number</i> [verbose]	
Display routing information permitted by an IPv6 prefix list	display ipv6 routing-table ipv6-prefix <i>ipv6-prefix-name</i> [verbose]	
Display IPv6 routing information of a routing protocol	display ipv6 routing-table protocol <i>protocol</i> [inactive verbose]	
Display IPv6 routing statistics	display ipv6 routing-table statistics	
Display IPv6 routing information for an IPv6 address range	display ipv6 routing-table <i>ipv6-address1</i> <i>prefix-length1</i> <i>ipv6-address2</i> <i>prefix-length2</i> [verbose]	
Clear specified IPv6 routing table statistics	reset ipv6 routing-table statistics protocol { all <i>protocol</i> }	Available in user view

25

GR OVERVIEW

Go to these sections for information you are interested in:

- "Introduction to Graceful Restart" on page 247
- "Basic Concepts in Graceful Restart" on page 247
- "Graceful Restart Communication Procedure" on page 248
- "Graceful Restart Mechanism for Several Commonly Used Protocols" on page 250



Throughout this chapter, the term router and the router icon refers to a router in a generic sense or a Layer 3 switch running routing protocols.

Introduction to Graceful Restart

Graceful Restart ensures the continuity of packet forwarding when a routing protocol restarts.

The mechanism of Graceful Restart works as follows: after the routing protocol on a Graceful Restart capable device has restarted, the device will notify its neighbors to temporarily preserve its adjacency with them and the routing information. The neighbors will help the restarting device to update its routing information and to restore it to the state prior to the restart in minimal time. The routing and forwarding remain highly stable across the restart, the packet forwarding path remains the same, and the whole system can forward IP packets continuously. Hence, it is called "Graceful Restart".

Basic Concepts in Graceful Restart

A router with the Graceful Restart feature enabled is called a Graceful Restart capable router. It can perform a Graceful Restart when its routing protocol restarts. Routers that are not Graceful Restart capable will follow the normal restart procedures after a routing protocol restart.

- GR Restarter: Graceful restarting router, the router whose routing protocol has restarted due to administrator instructions or network failure. It must be Graceful Restart capable.
- GR Helper: The neighbor of the GR Restarter, which helps the GR Restarter to retain the routing information. It must be Graceful Restart capable.
- GR Session: A Graceful Restart session, which is the negotiation between the GR Restarter and the GR Helper. A GR session includes restart notification and communications across restart. Through this session, GR Restarter and GR Helper can know the GR capability of each other.
- GR Time: The time taken for the GR Restarter and the GR Helper to establish a session between them. Upon detection of the down state of a neighbor, the GR Helper will preserve the topology and routing information sent from the GR Restarter for a period as specified by the GR Time.

Graceful Restart Communication Procedure

Configure a device as GR Restarter in a network. This device and its GR Helper must support GR or be GR capable. Thus, when GR Restarter restarts, its GR Helper can know its restart process.

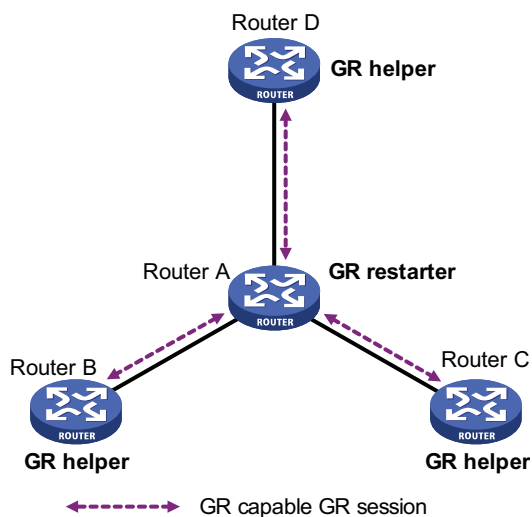


In some cases, GR Restarter and GR Helper can replace with each other.

The communication procedure between the GR Restarter and the GR Helper works as follows:

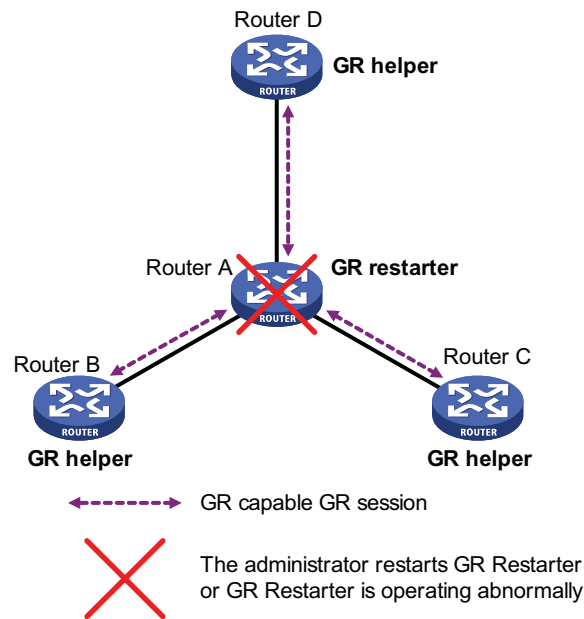
- 1 A GR session is established between the GR Restarter and the GR Helper.

Figure 64 A GR session is established between the GR Restarter and the GR Helper



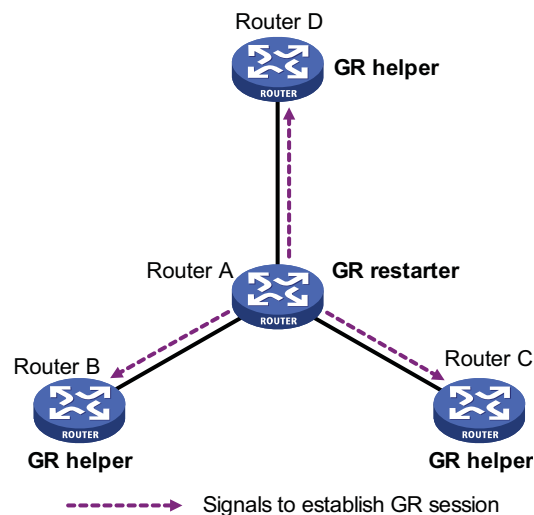
As illustrated in Figure 64, Router A works as GR Restarter, Router B, Router C and Router D are the GR Helpers of Router A. A GR session is established between the GR Restarter and the GR Helper.

- 2 GR Restarter restarting

Figure 65 Restarting process for the GR Restarter

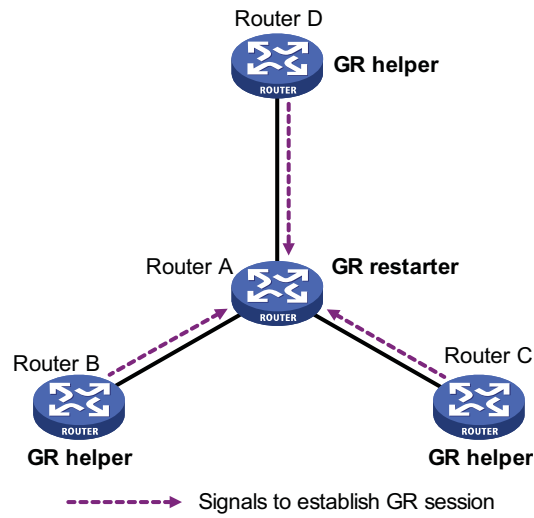
As illustrated in Figure 65. The GR Helper detects that the GR Restarter has restarted its routing protocol and assumes that it will recover within the GR Time. Before the GR Time expires, the GR Helper will neither terminate the session with the GR Restarter nor delete the topology or routing information of the latter.

3 GR Restarter signaling to GR Helper

Figure 66 The GR Restarter signals to the GR Helper(s) after restart

As illustrated in Figure 66, after the GR Restarter has recovered, it will signal to all its neighbors and will reestablish GR Session.

4 The GR Restarter obtaining topology and routing information from the GR Helper

Figure 67 The GR Restarter obtains topology and routing information from the GR Helper

As illustrated in Figure 67, the GR Restarter obtains the necessary topology and routing information from all its neighbors through the GR sessions between them and calculates its own routing table based on this information.

Graceful Restart Mechanism for Several Commonly Used Protocols

The switch supports Graceful Restart based on Border Gateway Protocol (BGP), Open Shortest Path First (OSPF), and Intermediate System to Intermediate System (IS-IS).

For the implementation and configuration procedure of the Graceful Restart mechanism of the above protocols, refer to "BGP Configuration" on page 365, "OSPF Configuration" on page 273, and "IS-IS Configuration" on page 325.

26

STATIC ROUTING CONFIGURATION

When configuring a static route, go to these sections for information you are interested in:

- "Introduction" on page 251
- "Configuring a Static Route" on page 252
- "Application Environment of Static Routing" on page 252
- "Displaying and Maintaining Static Routes" on page 254
- "Configuration Example" on page 254



The term "router" in this document refers to a router in a generic sense or a Layer 3 switch.

Introduction

Static Route A static route is a special route that is manually configured by the network administrator. If a network's topology is simple, you only need to configure static routes for the network to work normally. The proper configuration and usage of static routes can improve network performance and ensure bandwidth for important network applications.

The disadvantage of using static routes is that they cannot adapt to network topology changes. If a fault or a topological change occurs in the network, the routes will be unreachable and the network breaks. In this case, the network administrator has to modify the static routes manually.

Default Route A router selects the default route only when it cannot find any matching entry in the routing table.

If the destination address of a packet fails to match any entry in the routing table, the router selects the default route to forward the packet.

If there is no default route and the destination address of the packet fails to match any entry in the routing table, the packet will be discarded and an ICMP packet will be sent to the source to report that the destination or the network is unreachable.

You can create the default route with both destination and mask being 0.0.0.0, and some dynamic routing protocols, such as OSPF, RIP and IS-IS, can also generate the default route.

Application Environment of Static Routing

Before configuring a static route, you need to know the following concepts:

1 Destination address and mask

In the **ip route-static** command, an IPv4 address is in dotted decimal format and a mask can be either in dotted decimal format or in the form of mask length (the digits of consecutive 1s in the mask).

2 Output interface and next hop address

While configuring a static route, you can specify either the output interface or the next hop address depending on the specific occasion. The next hop address can not be a local interface IP address; otherwise, the route configuration will not take effect.

In fact, all the route entries must have a next hop address. When forwarding a packet, a router first searches the routing table for the route to the destination address of the packet. The system can find the corresponding link layer address and forward the packet only after the next hop address is specified.

When specifying the output interface, note that:

- If the output interface is a NULL 0 interface, there is no need to configure the next hop address.
- You are not recommended to specify a broadcast interface (such as VLAN interface) as the output interface, because a broadcast interface may have multiple next hops. If you have to do so, you must specify the corresponding next hop for the output interface.
- Other attributes

You can configure different preferences for different static routes so that route management policies can be applied more flexibly. For example, specifying the same preference for different routes to the same destination enables load sharing, while specifying different preferences for these routes enables route backup.

Configuring a Static Route

Configuration Prerequisites

Before configuring a static route, you need to configure the IP addresses for related interfaces.

Configuration Procedure

Follow these steps to configure a static route:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Configure a static route	ip route-static <i>dest-address</i> { <i>mask</i> <i>mask-length</i> } { <i>next-hop-address</i> <i>interface-type interface-number</i> [<i>next-hop-address</i>] } [preference <i>preference-value</i>] [tag <i>tag-value</i>] [description <i>description-text</i>]	Required By default, preference for static routes is 60, tag is 0, and no description information is configured.
Configure the default preference for static routes	ip route-static default-preference <i>default-preference-value</i>	Optional 60 by default



- *When configuring a static route, the static route does not take effect if you specify the next hop address first and then configure it as the IP address of a local interface, such as a VLAN interface.*
- *If you do not specify the preference when configuring a static route, the default preference will be used. Reconfiguring the default preference applies only to newly created static routes.*
- *You can flexibly control static routes by configuring tag values and using the tag values in the routing policy.*
- *If the destination IP address and mask are both configured as 0.0.0.0 with the **ip route-static** command, the route is the default route.*

Detecting Reachability of the Static Route's Nexthop

If a static route fails due to a topology change or a fault, the connection will be interrupted. To improve network stability, the system needs to detect reachability of the static route's next hop and switch to a backup route once the next hop is unreachable.

Detecting Nexthop Reachability Through Track

If you specify the nexthop but not outgoing interface when configuring a static route, you can associate the static route with a track entry to check the static route validity:

- When the track entry is positive, the static route's nexthop is reachable and the static route takes effect.
- When the track entry is negative, the static route's nexthop is unreachable and the static route is invalid. For details about track, refer to "Track Configuration" on page 1237.

Network requirements

To detect the reachability of a static route's nexthop through a Track entry, you need to create a Track first. For detailed Track configuration procedure, refer to "Track Configuration Task List" on page 1238.

Configuration procedure

Follow these steps to detect the reachability of a static route's nexthop through Track:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Associate the static route with a track entry	ip route-static <i>dest-address</i> { <i>mask</i> <i>mask-length</i> } <i>next-hop-address</i> track <i>track-entry-number</i> [preference <i>preference-value</i>] [tag <i>tag-value</i>] [description <i>description-text</i>]	Required Not configured by default



- To configure this feature for an existing static route, simply associate the static route with a track entry. For a non-existent static route, configure it and associate it with a Track entry.
- If a static route needs route recursion, the associated track entry must monitor the nexthop of the recursive route instead of that of the static route; otherwise, a valid route may be mistakenly considered invalid.

Displaying and Maintaining Static Routes

To do...	Use the command...	Remarks
Display the current configuration information	display current-configuration	Available in any view
Display the brief information of the IP routing table	display ip routing-table	
Display the detailed information of the IP routing table	display ip routing-table verbose	
View information of static routes	display ip routing-table protocol static [inactive verbose]	
Delete all the static routes	delete static-routes all	Available In system view

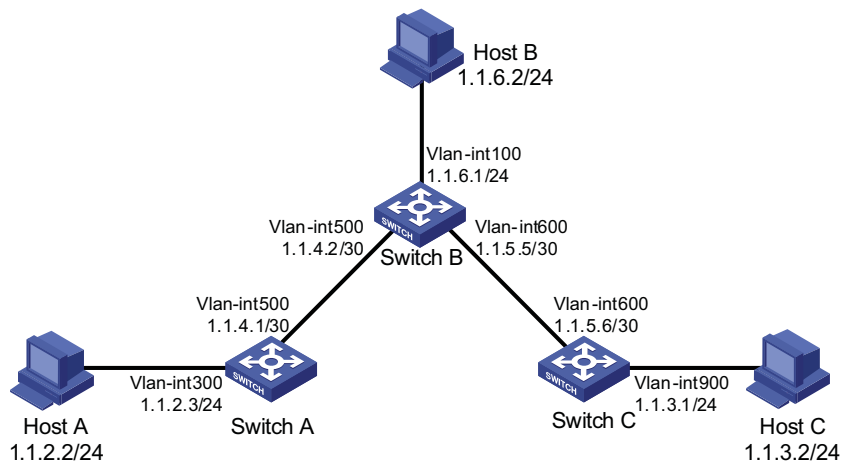
Configuration Example

Network requirements

The IP addresses and masks of the switches and hosts are shown in the following figure. Static routes are required for interconnection between any two hosts.

Network diagram

Figure 68 Network diagram for static route configuration



Configuration procedure

- 1 Configuring IP addresses for interfaces (omitted)
- 2 Configuring static routes

Configure a default route on Switch A

```
<SwitchA> system-view
[SwitchA] ip route-static 0.0.0.0 0.0.0.0 1.1.4.2
```

Configure two static routes on Switch B

```
<SwitchB> system-view
[SwitchB] ip route-static 1.1.2.0 255.255.255.0 1.1.4.1
[SwitchB] ip route-static 1.1.3.0 255.255.255.0 1.1.5.6
```

Configure a default route on Switch C

```
<SwitchC> system-view
[SwitchC] ip route-static 0.0.0.0 0.0.0.0 1.1.5.5
```

- 1 Configure the hosts

The default gateways for the three hosts A, B and C are 1.1.2.3, 1.1.6.1 and 1.1.3.1 respectively. The configuration procedure is omitted.

- 1 Display the configuration result

Display the IP routing table of Switch A.

```
[SwitchA] display ip routing-table
Routing Tables: Public
    Destinations : 7          Routes : 7
```

Destination/Mask	Proto	Pre	Cost	NextHop	Interface
0.0.0.0/0	Static	60	0	1.1.4.2	Vlan500
1.1.2.0/24	Direct	0	0	1.1.2.3	Vlan300
1.1.2.3/32	Direct	0	0	127.0.0.1	InLoop0
1.1.4.0/30	Direct	0	0	1.1.4.1	Vlan500
1.1.4.1/32	Direct	0	0	127.0.0.1	InLoop0
127.0.0.0/8	Direct	0	0	127.0.0.1	InLoop0
127.0.0.1/32	Direct	0	0	127.0.0.1	InLoop0

Display the IP routing table of Switch B.

```
[SwitchB] display ip routing-table
Routing Tables: Public
    Destinations : 10         Routes : 10
```

Destination/Mask	Proto	Pre	Cost	NextHop	Interface
1.1.2.0/24	Static	60	0	1.1.4.1	Vlan500
1.1.3.0/24	Static	60	0	1.1.5.6	Vlan600
1.1.4.0/30	Direct	0	0	1.1.4.2	Vlan500
1.1.4.2/32	Direct	0	0	127.0.0.1	InLoop0
1.1.5.0/30	Direct	0	0	1.1.5.5	Vlan600
1.1.5.5/32	Direct	0	0	127.0.0.1	InLoop0
127.0.0.0/8	Direct	0	0	127.0.0.1	InLoop0
127.0.0.1/32	Direct	0	0	127.0.0.1	InLoop0
1.1.6.0/24	Direct	0	0	1.1.6.1	Vlan100
1.1.6.1/32	Direct	0	0	127.0.0.1	InLoop0

From Host A, use the **ping** command to verify the network layer reachability to Host B and Host C.

27

RIP CONFIGURATION



- *The term “router” in this document refers to a router in a generic sense or a Layer 3 switch.*
- *The Switch 4800G only support single RIP process.*

When configuring RIP, go to these sections for information you are interested in:

- “RIP Overview” on page 257
- “Configuring RIP Basic Functions” on page 261
- “Configuring RIP Route Control” on page 263
- “Configuring RIP Network Optimization” on page 266
- “Displaying and Maintaining RIP” on page 269
- “RIP Configuration Examples” on page 269
- “Troubleshooting RIP” on page 271

RIP Overview

RIP is a simple Interior Gateway Protocol (IGP), mainly used in small-sized networks, such as academic networks and simple LANs. RIP is not applicable to complex networks.

RIP is still widely used in practical networking due to easier implementation, configuration and maintenance than OSPF and IS-IS.

RIP Working Mechanism **Basic concepts**

RIP is a distance vector routing protocol, using UDP packets for exchanging information through port 520.

RIP uses a hop count to measure the distance to a destination. The hop count is known as the metric. The hop count from a router to a directly connected network is 0. The hop count from one router to a directly connected router is 1. To limit convergence time, the range of RIP metric value is from 0 to 15. A metric value of 16 (or bigger) is considered infinite, which means the destination network is unreachable. That is why RIP is not suitable for large-scaled networks.

RIP prevents routing loops by implementing the split horizon and poison reverse functions.

RIP routing table

A RIP router has a routing table containing routing entries of all reachable destinations, and each routing entry contains:

- Destination address: IP address of a host or a network.

- Next hop: IP address of the adjacent router's interface to reach the destination.
- Egress interface: Packet outgoing interface.
- Metric: Cost from the local router to the destination.
- Route time: Time elapsed since the routing entry was last updated. The time is reset to 0 every time the routing entry is updated.
- Route tag: Identifies a route, used in a routing policy to flexibly control routes. For information about routing policy, refer to "Routing Policy Configuration" on page 415.

RIP timers

RIP employs four timers, update, timeout, suppress, and garbage-collect.

- The update timer defines the interval between routing updates.
- The timeout timer defines the route aging time. If no update for a route is received within the aging time, the metric of the route is set to 16 in the routing table.
- The suppress timer defines how long a RIP route stays in the suppressed state. When the metric of a route is 16, the route enters the suppressed state. In the suppressed state, only routes which come from the same neighbor and whose metric is less than 16 will be received by the router to replace unreachable routes.
- The garbage-collect timer defines the interval from when the metric of a route becomes 16 to when it is deleted from the routing table. During the garbage-collect timer length, RIP advertises the route with the routing metric set to 16. If no update is announced for that route after the garbage-collect timer expires, the route will be deleted from the routing table.

Routing loops prevention

RIP is a distance vector (D-V) routing protocol. Since a RIP router advertises its own routing table to neighbors, routing loops may occur.

RIP uses the following mechanisms to prevent routing loops.

- Counting to infinity. The metric value of 16 is defined as unreachable. When a routing loop occurs, the metric value of the route will increment to 16.
- Split horizon. A router does not send the routing information learned from a neighbor to the neighbor to prevent routing loops and save bandwidth.
- Poison reverse. A router sets the metric of routes received from a neighbor to 16 and sends back these routes to the neighbor to help delete useless information from the neighbor's routing table.
- Triggered updates. A router advertises updates once the metric of a route is changed rather than after the update period expires to speed up network convergence.

Operation of RIP The following procedure describes how RIP works.

- 1 After RIP is enabled, the router sends Request messages to neighboring routers. Neighboring routers return Response messages including information about their routing tables.
- 2 After receiving such information, the router updates its local routing table, and sends triggered update messages to its neighbors. All routers on the network do the same to keep the latest routing information.
- 3 By default, a RIP router sends its routing table to neighbors every 30 seconds.
- 4 RIP ages out routes by adopting an aging mechanism to keep only valid routes.

RIP Version RIP has two versions, RIPv1 and RIPv2.

RIPv1, a classful routing protocol, supports message advertisement via broadcast only. RIPv1 protocol messages do not carry mask information, which means it can only recognize routing information of natural networks such as Class A, B, C. That is why RIPv1 does not support discontinuous subnets.

RIPv2 is a classless routing protocol. Compared with RIPv1, RIPv2 has the following advantages.

- Supporting route tags. Route tags are used in routing policies to flexibly control routes.
- Supporting masks, route summarization and Classless Inter-Domain Routing (CIDR).
- Supporting designated next hops to select the best next hops on broadcast networks.
- Supporting multicast routing update to reduce resource consumption.
- Supporting plain text authentication and MD5 authentication to enhance security.

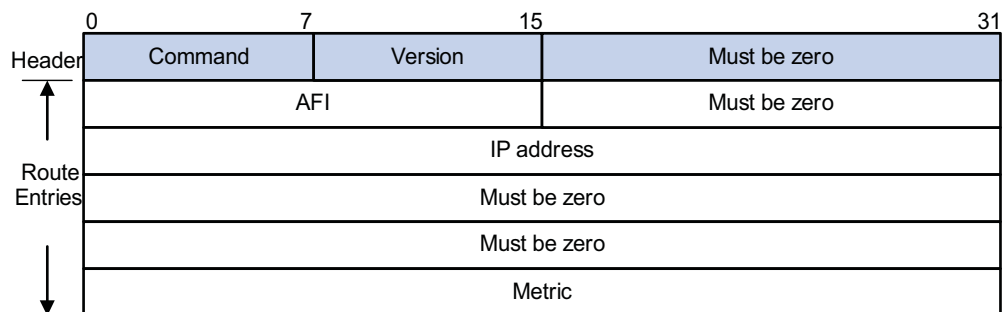


RIPv2 has two types of message transmission: broadcast and multicast. Multicast is the default type using 224.0.0.9 as the multicast address. The interface working in the RIPv2 broadcast mode can also receive RIPv1 messages.

RIP Message Format RIPv1 message format

A RIPv1 message consists of a header and up to 25 route entries.

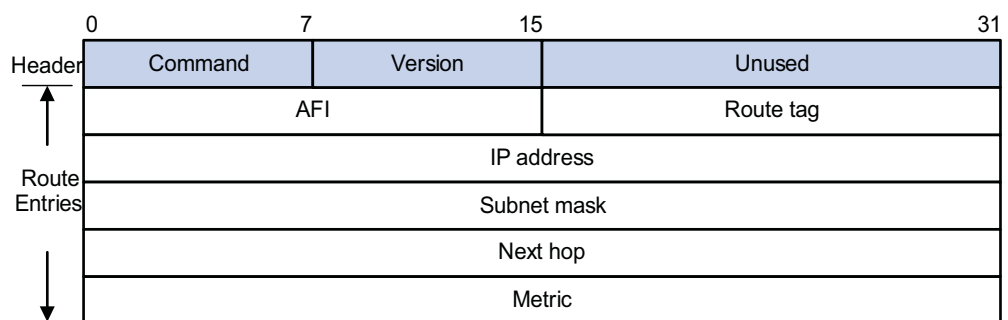
Figure 69 shows the format of RIPv1 message.

Figure 69 RIPv1 Message Format

- Command: Type of message. 1 indicates request, and 2 indicates response.
- Version: Version of RIP, 0x01 for RIPv1.
- AFI: Address Family Identifier, 2 for IP.
- IP Address: Destination IP address of the route. It can be a natural network, subnet or a host address.
- Metric: Cost of the route.

RIPv2 message format

The format of RIPv2 message is similar with RIPv1. Figure 70 shows it.

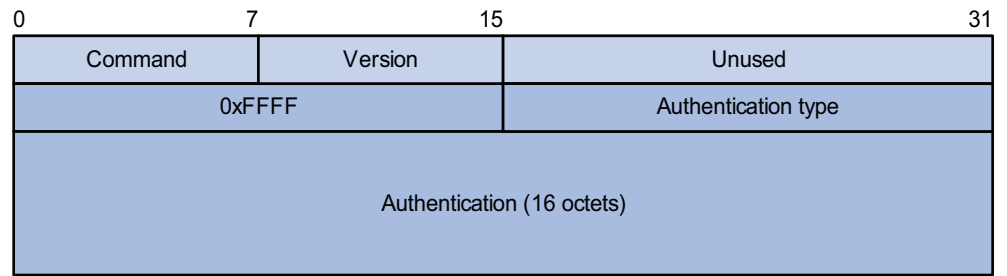
Figure 70 RIPv2 Message Format


The differences from RIPv1 are stated as following.

- Version: Version of RIP. For RIPv2 the value is 0x02.
- Route Tag: Route Tag.
- IP Address: Destination IP address. It could be a natural network address, subnet address or host address.
- Subnet Mask: Mask of the destination address.
- Next Hop: If set to 0.0.0.0, it indicates that the originator of the route is the best next hop; otherwise it indicates a next hop better than the originator of the route.

RIPv2 authentication

RIPv2 sets the AFI field of the first route entry to 0xFFFF to identify authentication information. See Figure 71.

Figure 71 RIPv2 Authentication Message

- Authentication Type: 2 represents plain text authentication, while 3 represents MD5.
 - Authentication: Authentication data, including password information when plain text authentication is adopted or including key ID, MD5 authentication data length and sequence number when MD5 authentication is adopted.
-  ■ *RFC 1723 only defines plain text authentication. For information about MD5 authentication, refer to RFC2082 "RIPv2 MD5 Authentication".*
- *With RIPv1, you can configure the authentication mode in interface view. However, the configuration will not take effect because RIPv1 does not support authentication.*

Supported RIP Features The current implementation supports RIPv1 and RIPv2

Protocols and Standards

- RFC 1058: Routing Information Protocol
- RFC 1723: RIP Version 2 - Carrying Additional Information
- RFC 1721: RIP Version 2 Protocol Analysis
- RFC 1722: RIP Version 2 Protocol Applicability Statement
- RFC 1724: RIP Version 2 MIB Extension
- RFC 2082: RIPv2 MD5 Authentication

Configuring RIP Basic Functions

Configuration Prerequisites Before configuring RIP basic functions, configure IP addresses for interfaces, making all adjacent nodes reachable to each other at the network layer.

Configuration Procedure **Enabling RIP and a RIP interface**

Follow these steps to enable RIP:

To do...	Use the command...	Remarks
Enter system view	System-view	--
Enable a RIP process and enter RIP view	rip [process-id]	Required Not enabled by default

To do...	Use the command...	Remarks
Enable RIP on the interface attached to the specified network	network <i>network-address</i>	Required Disabled by default



- If you make some RIP configurations in interface view before enabling RIP, those configurations will take effect after RIP is enabled.
- RIP runs only on the interfaces residing on the specified networks. Therefore, you need to specify the network after enabling RIP to validate RIP on a specific interface.
- You can enable RIP on all interfaces using the command **network 0.0.0.0**.

Configuring the interface behavior

Follow these steps to configure the interface behavior:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIP view	rip [<i>process-id</i>]	--
Disable an or all interfaces from sending routing updates (the interfaces can still receive updates)	silent-interface { all <i>interface-type</i> <i>interface-number</i> }	Optional All interfaces can send routing updates by default.
Return to system view	quit	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable the interface to receive RIP messages	rip input	Optional Enabled by default
Enable the interface to send RIP messages	rip output	Optional Enabled by default

Configuring a RIP version

You can configure a RIP version in RIP or interface view.

- If neither global nor interface RIP version is configured, the interface sends RIPv1 broadcasts and can receive RIPv1 broadcast and unicast packets, and RIPv2 broadcast, multicast, and unicast packets.
- If an interface has no RIP version configured, it uses the global RIP version; otherwise it uses the RIP version configured on it.
- With RIPv1 configured, an interface sends RIPv1 broadcasts, and can receive RIPv1 broadcasts and RIPv1 unicasts.
- With RIPv2 configured, a multicast interface sends RIPv2 multicasts and can receive RIPv2 unicasts, broadcasts and multicasts.
- With RIPv2 configured, a broadcast interface sends RIPv2 broadcasts and can receive RIPv1 unicasts, and broadcasts, and RIPv2 broadcasts, multicasts and unicasts.

Follow these steps to configure a RIP version:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIP view	rip [<i>process-id</i>]	--
Specify a global RIP version	version { 1 2 }	Optional By default, if an interface has a RIP version specified, the version takes precedence over the global one. If no RIP version is specified for an interface, the interface can send RIPv1 broadcasts, unicasts, RIPv2 broadcasts, multicasts and unicasts.
Return to system view	Quit	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--
Specify a RIP version for the interface	rip version { 1 2 [broadcast multicast] }	Optional

Configuring RIP Route Control

In complex networks, you need to configure advanced RIP features.

This section covers the following topics:

- “Configuring an Additional Routing Metric” on page 263
- “Configuring RIPv2 Route Summarization” on page 264
- “Disabling Host Route Reception” on page 264
- “Advertising a Default Route” on page 265
- “Configuring Inbound/Outbound Route Filtering” on page 265
- “Configuring a Priority for RIP” on page 266
- “Configuring RIP Route Redistribution” on page 266

Before configuring RIP routing feature, complete the following tasks:

- Configure an IP address for each interface, and make sure all neighboring routers are reachable to each other.
- Configure RIP basic functions

Configuring an Additional Routing Metric

An additional routing metric can be added to the metric of an inbound or outbound RIP route.

The outbound additional metric is added to the metric of a sent route, the route’s metric in the routing table is not changed.

The inbound additional metric is added to the metric of a received route before the route is added into the routing table, so the route’s metric is changed.

Follow these steps to configure additional routing metrics:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--
Define an inbound additional routing metric	rip metricin [route-policy <i>route-policy-name</i>] <i>value</i>	Optional 0 by default
Define an outbound additional routing metric	rip metricout [route-policy <i>route-policy-name</i>] <i>value</i>	Optional 1 by default

Configuring RIPv2 Route Summarization

Route summarization means that subnets in a natural network are summarized with a natural network that is sent to other networks. This feature can reduce the size of routing tables.

Enabling RIPv2 route automatic summarization

You can disable RIPv2 route automatic summarization if you want to advertise all subnet routes.

Follow these steps to enable RIPv2 route automatic summarization:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIP view	rip [<i>process-id</i>]	--
Enable RIPv2 automatic route summarization	summary	Optional Enabled by default

Advertising a summary route

You can configure RIPv2 to advertise a summary route on the specified interface.

To do so, use the following commands:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIP view	rip [<i>process-id</i>]	--
Disable RIPv2 automatic route summarization	undo summary	Required Enabled by default
Return to system view	quit	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Advertise a summary route	rip summary-address <i>ip-address</i> { <i>mask</i> <i>mask-length</i> }	Required



You need to disable RIPv2 route automatic summarization before advertising a summary route on an interface.

Disabling Host Route Reception

Sometimes a router may receive many host routes from the same network, which are not helpful for routing and occupy a large amount of network resources. In this case, you can disable RIP from receiving host routes to save network resources.

Follow these steps to disable RIP from receiving host routes:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter RIP view	rip [<i>process-id</i>]	-
Disable RIP from receiving host routes	undo host-route	Required Enabled by default



RIPv2 can be disabled from receiving host routes, but RIPv1 cannot.

Advertising a Default Route

You can configure RIP to advertise a default route with A specified metric to RIP neighbors.

Follow these steps to configure RIP to advertise a default route:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIP view	rip [<i>process-id</i>]	--
Enable RIP to advertise a default route	default-route originate cost <i>value</i>	Required Not enabled by default



The router enabled to advertise a default route does not receive default routes from RIP neighbors.

Configuring Inbound/Outbound Route Filtering

The device supports route filtering. You can filter routes by configuring the inbound and outbound route filtering policies via referencing an ACL or IP prefix list. You can also configure the router to receive only routes from a specified neighbor.

Follow these steps to configure route filtering:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIP view	rip [<i>process-id</i>]	--
Configure the filtering of incoming routes	filter-policy { <i>acl-number</i> gateway <i>ip-prefix-name</i> ip-prefix <i>ip-prefix-name</i> [gateway <i>ip-prefix-name</i>] } import [<i>interface-type</i> <i>interface-number</i>]	Required Not configured by default
Configure the filtering of outgoing routes	filter-policy { <i>acl-number</i> ip-prefix <i>ip-prefix-name</i> } export [<i>protocol</i> [<i>process-id</i>] <i>interface-type</i> <i>interface-number</i>]	Required Not configured by default



- Using the **filter-policy import** command filters incoming routes. Routes not passing the filtering will be neither installed into the routing table nor advertised to neighbors.
- Using the **filter-policy export** command filters outgoing routes, including routes redistributed with the **import-route** command.

Configuring a Priority for RIP

Multiple IGP protocols may run in a router. If you want RIP routes to have a higher priority than those learned by other routing protocols, you can assign RIP a smaller priority value to influence optimal route selection.

Follow these steps to configure a priority for RIP:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIP view	rip [<i>process-id</i>]	--
Configure a priority for RIP	preference [route-policy <i>route-policy-name</i>] <i>value</i>	Optional 100 by default

Configuring RIP Route Redistribution

Follow these steps to configure RIP route redistribution:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIP view	rip [<i>process-id</i>]	--
Configure a default metric for redistributed routes	default-cost <i>value</i>	Optional The default metric of a redistributed route is 0 by default.
Redistribute routes from another protocol	import-route <i>protocol</i> [<i>process-id</i>] [allow-ibgp] [cost <i>cost</i> route-policy <i>route-policy-name</i> tag <i>tag</i>] *	Required No redistribution is configured by default.

Configuring RIP Network Optimization

Complete the following tasks before configuring RIP network optimization:

- Configure network addresses for interfaces, and make neighboring nodes reachable to each other;
- Configure RIP basic functions.

Configuring RIP Timers

Follow these steps to configure RIP timers:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIP view	rip [<i>process-id</i>]	--
Configure values for RIP timers	timers { garbage-collect <i>garbage-collect-value</i> suppress <i>suppress-value</i> timeout <i>timeout-value</i> update <i>update-value</i> }*	Optional The default update timer, timeout timer, suppress timer, and garbage-collect timer are 30s, 180s, 120s and 120s respectively.



Based on network performance, you need to make RIP timers of RIP routers identical to each other to avoid unnecessary traffic or route oscillation.

Configuring Split Horizon and Poison Reverse



If both split horizon and poison reverse are configured, only the poison reverse function takes effect.

Enabling split horizon

The split horizon function disables an interface from sending routes received from the interface to prevent routing loops between adjacent routers.

Follow these steps to enable split horizon:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type interface-number</i>	-
Enable split horizon	rip split-horizon	Optional Enabled by default



Disabling the split horizon function on a point-to-point link does not take effect.

Enabling poison reverse

The poison reverse function allows an interface to advertise the routes received from it, but the metric of these routes is set to 16, making them unreachable.

Follow these steps to enable poison reverse:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type interface-number</i>	-
Enable poison reverse	rip poison-reverse	Required Disabled by default

Configuring the Maximum Number of Load Balanced Routes

Follow these steps to configure the maximum number of load balanced routes:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIP view	rip [<i>process-id</i>]	--
Configure the maximum number of load balanced routes	maximum load-balancing <i>number</i>	Optional The default maximum number is 4.

Enabling Zero Field Check on Incoming RIPv1 Messages

Some fields in the RIPv1 message must be zero. These fields are called zero fields. You can enable zero field check on received RIPv1 messages. If such a field contains a non-zero value, the RIPv1 message will not be processed. If you are sure

that all messages are trustworthy, you can disable zero field check to save CPU resources.

Follow these steps to enable zero field check on incoming RIPv1 messages:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIP view	rip [<i>process-id</i>]	--
Enable zero field check on received RIPv1 messages	checkzero	Optional Enabled by default

Enabling Source IP Address Check on Incoming RIP Updates

You can enable source IP address check on incoming RIP updates.

For a message received on an Ethernet interface, RIP compares the source IP address of the message with the IP address of the interface. If they are not in the same network segment, RIP discards the message.

For a message received on a serial interface, RIP checks whether the source address of the message is the IP address of the peer interface. If not, RIP discards the message.

Follow these steps to enable source IP address check on incoming RIP updates:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIP view	rip [<i>process-id</i>]	--
Enable source IP address check on incoming RIP messages	validate-source-address	Optional Enabled by default



The source IP address check feature should be disabled if a RIP neighbor is not directly connected.

Configuring RIPv2 Message Authentication

RIPv2 supports two authentication modes: plain text and MD5.

In plain text authentication, the authentication information is sent with the RIP message, which however cannot meet high security needs.

Follow these steps to configure RIPv2 message authentication:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter interface view	interface <i>interface-type interface-number</i>	--
Configure RIPv2 authentication	rip authentication-mode { md5 { rfc2082 <i>key-string key-id</i> rfc2453 <i>key-string</i> } simple <i>password</i> }	Required

Specifying a RIP Neighbor

Usually, RIP sends messages to broadcast or multicast addresses. On non broadcast or multicast links, you need to manually specify RIP neighbors. If a specified

neighbor is not directly connected, you must disable source address check on incoming updates.

Follow these steps to specify a RIP neighbor:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIP view	rip [<i>process-id</i>]	--
Specify a RIP neighbor	peer <i>ip-address</i>	Required By default, RIP sends no updates to any IP address.
Disable source address check on incoming RIP updates	undo validate-source-address	Required Not disabled by default



You need not use the **peer** *ip-address* command when the neighbor is directly connected; otherwise the neighbor may receive both the unicast and multicast (or broadcast) of the same routing information.

Displaying and Maintaining RIP

To do...	Use the command...	Remarks
Display RIP current status and configuration information	display rip [<i>process-id</i>]	Available in any view
Display all active routes in RIP database	display rip <i>process-id</i> database	
Display RIP interface information	display rip <i>process-id</i> interface [<i>interface-type</i> <i>interface-number</i>]	
Display routing information about a specified RIP process	display rip <i>process-id</i> route [statistics <i>ip-address</i> { <i>mask</i> <i>mask-length</i> } peer <i>ip-address</i>]	
Clear the statistics of a RIP process	reset rip <i>process-id</i> statistics	Available in user view

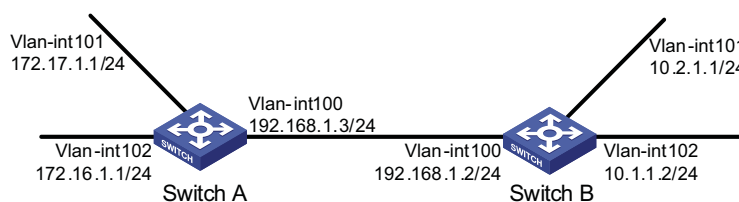
RIP Configuration Examples

Configuring RIP Version Network requirements

As shown in Figure 72, enable RIPv2 on all interfaces on Switch A and Switch B.

Network diagram

Figure 72 Network diagram for RIP version configuration



Configuration procedure

- 1 Configure IP addresses for interfaces (omitted).
- 2 Configure basic RIP functions

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] rip
[SwitchA-rip-1] network 192.168.1.0
[SwitchA-rip-1] network 172.16.0.0
[SwitchA-rip-1] network 172.17.0.0
[SwitchA-rip-1] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] rip
[SwitchB-rip-1] network 192.168.1.0
[SwitchB-rip-1] network 10.0.0.0
[SwitchB-rip-1] quit
```

Display the RIP routing table of Switch A.

```
[SwitchA] display rip 1 route
Route Flags: R - RIP, T - TRIP
              P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
-----
Peer 192.168.1.2 on Vlan-interface100
  Destination/Mask    Nexthop    Cost    Tag    Flags    Sec
  10.0.0.0/8          192.168.1.2  1       0     RA       11
```

From the routing table, you can find RIPv1 uses natural mask.

- 3 Configure RIP version

Configure RIPv2 on Switch A.

```
[SwitchA] rip
[SwitchA-rip-1] version 2
[SwitchA-rip-1] undo summary
```

Configure RIPv2 on Switch B.

```
[SwitchB] rip
[SwitchB-rip-1] version 2
[SwitchB-rip-1] undo summary
```

Display the RIP routing table on Switch A.

```
[SwitchA] display rip 1 route
Route Flags: R - RIP, T - TRIP
              P - Permanent, A - Aging, S - Suppressed, G - Garbage-collect
-----
Peer 192.168.1.2 on Vlan-interface100
  Destination/Mask    Nexthop    Cost    Tag    Flags    Sec
  10.2.1.0/24         192.168.1.2  1       0     RA       16
  10.1.1.0/24         192.168.1.2  1       0     RA       16
```

From the routing table, you can see RIPv2 uses classless subnet masks.



Since RIPv1 routing information has a long aging time, it will still exist until aged out after RIPv2 is configured.

Troubleshooting RIP

No RIP Updates Received **Symptom:**

No RIP updates are received when the links work well.

Analysis:

After enabling RIP, you must use the **network** command to enable corresponding interfaces. Make sure no interfaces are disabled from handling RIP messages.

If the peer is configured to send multicast messages, the same should be configured on the local end.

Solution:

- Use the **display current-configuration** command to check RIP configuration
- Use the **display rip** command to check whether some interface is disabled

Route Oscillation Occurred **Symptom:**

When all links work well, route oscillation occurs on the RIP network. After displaying the routing table, you may find some routes appear and disappear in the routing table intermittently.

Analysis:

In the RIP network, make sure all the same timers within the whole network are identical and relationships between timers are reasonable. For example, the timeout timer value should be larger than the update timer value.

Solution:

- Use the **display rip** command to check the configuration of RIP timers
- Use the **timers** command to adjust timers properly.

28

OSPF CONFIGURATION



The term “router” in this document refers to a router in a generic sense or a Layer 3 switch.

Open Shortest Path First (OSPF) is a link state interior gateway protocol developed by the OSPF working group of the Internet Engineering Task Force (IETF). At present, OSPF version 2 (RFC2328) is used.

When configuring OSPF, go to these sections for information you are interested in:

- “Introduction to OSPF” on page 273
- “OSPF Configuration Task List” on page 292
- “Configuring OSPF Basic Functions” on page 293
- “Configuring OSPF Area Parameters” on page 294
- “Configuring OSPF Network Types” on page 295
- “Configuring OSPF Route Control” on page 297
- “Configuring OSPF Network Optimization” on page 300
- “Configuring OSPF Graceful Restart” on page 306
- “Displaying and Maintaining OSPF” on page 309
- “OSPF Configuration Examples” on page 309
- “Troubleshooting OSPF Configuration” on page 323

Introduction to OSPF



Unless otherwise noted, OSPF refers to OSPFv2 throughout this document.

OSPF has the following features:

- Wide scope: Supports networks of various sizes and up to several hundred routers in an OSPF routing domain.
- Fast convergence: Transmits updates instantly after network topology changes for routing information synchronization in the AS.
- Loop-free: Computes routes with the shortest path first (SPF) algorithm according to the collected link states, so no route loops are generated.
- Area partition: Allows an AS to be split into different areas for ease of management and the routing information transmitted between areas is summarized to reduce network bandwidth consumption.
- Equal-cost multi-route: Supports multiple equal-cost routes to a destination.

- Routing hierarchy: Supports a four-level routing hierarchy that prioritizes the routes into intra-area, inter-area, external Type-1, and external Type-2 routes.
- Authentication: Supports interface-based packet authentication to guarantee the security of packet exchange.
- Multicast: Supports packet multicasting on some types of links.

Basic Concepts **Autonomous System**

A set of routers using the same routing protocol to exchange routing information constitute an Autonomous System (AS).

OSPF route computation

OSPF route computation is described as follows:

- Based on the network topology around itself, each router generates Link State Advertisements (LSA) and sends them to other routers in update packets.
- Each OSPF router collects LSAs from other routers to compose a LSDB (Link State Database). An LSA describes the network topology around a router, so the LSDB describes the entire network topology of the AS.
- Each router transforms the LSDB to a weighted directed graph, which actually reflects the topology architecture of the entire network. All the routers have the same graph.
- Each router uses the SPF algorithm to compute a Shortest Path Tree that shows the routes to the nodes in the autonomous system. The router itself is the root of the tree.

Router ID

To run OSPF, a router must have a Router ID, which is a 32-bit unsigned integer, the unique identifier of the router in the AS.

You may assign a Router ID to an OSPF router manually. If no Router ID is specified, the system automatically selects one for the router as follows:

- If the loopback interfaces are configured, select the highest IP address among them.
- If no loopback interface is configured, select the highest IP address among addresses of active interfaces on the router.

OSPF packets

OSPF uses five types of packets:

- Hello packet: Periodically sent to find and maintain neighbors, containing the values of some timers, information about the DR, BDR and known neighbors.
- DD packet (database description packet): Describes the digest of each LSA in the LSDB, exchanged between two routers for data synchronization.
- LSR (link state request) packet: Requests needed LSAs from the neighbor. After exchanging the DD packets, the two routers know which LSAs of the neighbor are missing from the local LSDBs. In this case, they send an LSR packet to each other, requesting the missing LSAs. The LSA packet contains the digest of the missing LSAs.

- LSU (link state update) packet: Transmits the needed LSAs to the neighbor.
- LSAck (link state acknowledgment) packet: Acknowledges received LSU packets. It contains the headers of received LSAs (a packet can acknowledge multiple LSAs).

LSA types

OSPF sends routing information in LSAs, which, as defined in RFC 2328, have the following types:

- Router LSA: Type-1 LSA, originated by all routers, flooded throughout a single area only. This LSA describes the collected states of the router's interfaces to an area.
- Network LSA: Type-2 LSA, originated for broadcast and NBMA networks by the designated router, flooded throughout a single area only. This LSA contains the list of routers connected to the network.
- Network Summary LSA: Type-3 LSA, originated by ABRs (Area Border Routers), and flooded throughout the LSA's associated area. Each summary-LSA describes a route to a destination outside the area, yet still inside the AS (an inter-area route).
- ASBR Summary LSA: Type-4 LSA, originated by ABRs and flooded throughout the LSA's associated area. Type 4 summary-LSAs describe routes to ASBR (Autonomous System Boundary Router).
- AS External LSA: Type-5 LSA, originated by ASBRs, and flooded throughout the AS (except stub and NSSA areas). Each AS-external-LSA describes a route to another AS.
- NSSA External LSA: Type-7 LSA, as defined in RFC 1587, originated by ASBRs in NSSAs (Not-So-Stubby Areas) and flooded throughout a single NSSA. NSSA LSAs describe routes to other ASs.
- Opaque LSA: A proposed type of LSA, the format of which consists of a standard LSA header and application specific information. Opaque LSAs are used by the OSPF protocol or by some application to distribute information into the OSPF routing domain. The opaque LSA includes three types, Type 9, Type 10 and Type 11, which are used to flood into different areas. The Type 9 opaque LSA is flooded into the local subnet, the Type 10 is flooded into the local area, and the Type 11 is flooded throughout the whole AS.

Neighbor and Adjacency

In OSPF, the "Neighbor" and "Adjacency" are two different concepts.

Neighbor: Two routers that have interfaces to a common network. Neighbor relationships are maintained by, and usually dynamically discovered by, OSPF's hello packets. When a router starts, it sends a hello packet via the OSPF interface, and the router that receives the hello packet checks parameters carried in the packet. If parameters of the two routers match, they become neighbors.

Adjacency: A relationship formed between selected neighboring routers for the purpose of exchanging routing information. Not every pair of neighboring routers become adjacent, which depends on network types. Only by synchronizing the LSDB via exchanging DD packets and LSAs can two routers become adjacent.

OSPF Area Partition and Route Summarization

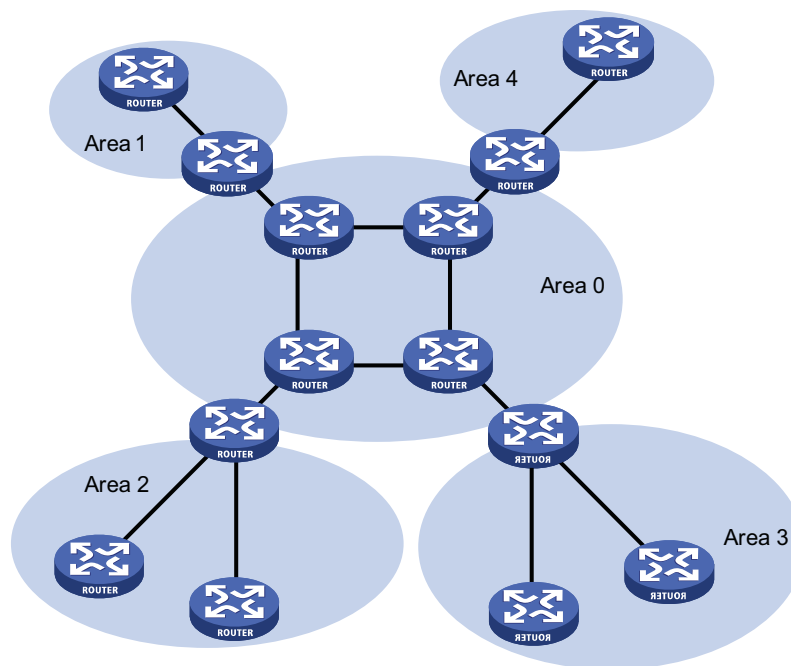
Area partition

When a large number of OSPF routers are present on a network, LSDBs may become so large that a great amount of storage space is occupied and CPU resources are exhausted by performing SPF computation.

In addition, as the topology of a large network is prone to changes, enormous OSPF packets may be created, reducing bandwidth utilization. Each topology change makes all routers perform route calculation.

To solve this problem, OSPF splits an AS into multiple areas, which are identified by area ID. The boundaries between areas are routers rather than links. A network segment (or a link) can only reside in one area, in other words, an OSPF interface must be specified to belong to its attached area, as shown in the figure below.

Figure 73 OSPF area partition



After area partition, area border routers perform route summarization to reduce the number of LSAs advertised to other areas and minimize the effect of topology changes.

Classification of Routers

The OSPF routers fall into four types according to the position in the AS:

1 Internal Router

All interfaces on an internal router belong to one OSPF area.

2 Area Border Router (ABR)

An area border router belongs to more than two areas, one of which must be the backbone area. It connects the backbone area to a non-backbone area. The connection between an area border router and the backbone area can be physical or logical.

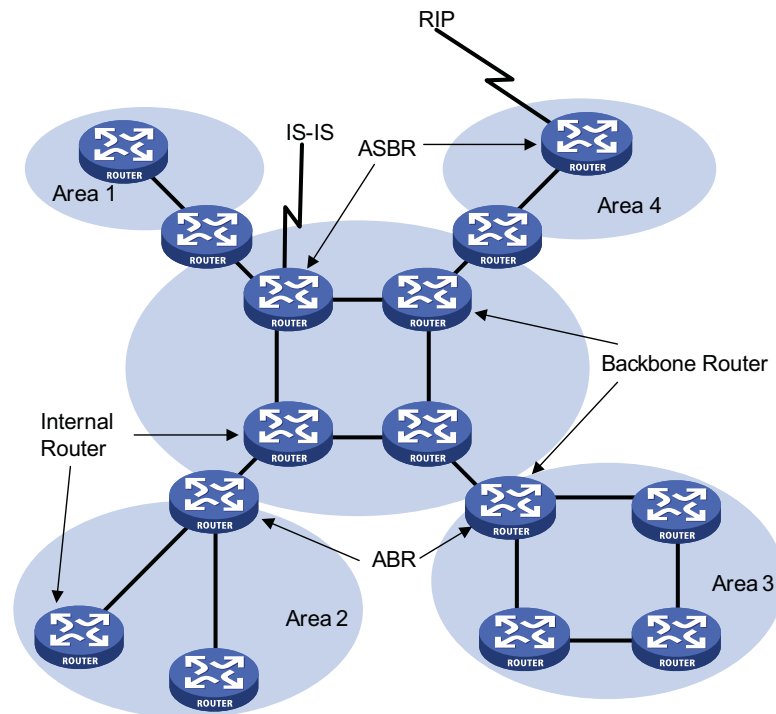
3 Backbone Router

At least one interface of a backbone router must be attached to the backbone area. Therefore, all ABRs and internal routers in area 0 are backbone routers.

4 Autonomous System Border Router (ASBR)

The router exchanging routing information with another AS is an ASBR, which may not reside on the boundary of the AS. It can be an internal router or area border router.

Figure 74 OSPF router types



Backbone area and virtual links

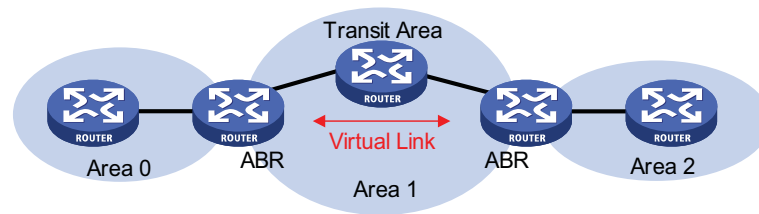
Each AS has a backbone area, which is responsible for distributing routing information between non-backbone areas. Routing information between non-backbone areas must be forwarded by the backbone area. Therefore, OSPF requires that:

- All non-backbone areas must maintain connectivity to the backbone area.
- The backbone area itself must maintain connectivity.

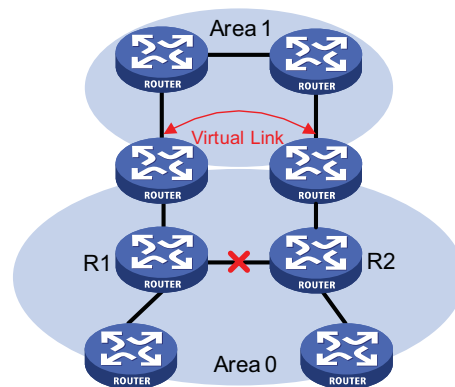
In practice, due to physical limitations, the requirements may not be satisfied. In this case, configuring OSPF virtual links is a solution.

A virtual link is established between two area border routers via a non-backbone area and is configured on both ABRs to take effect. The area that provides the non-backbone area internal route for the virtual link is a "transit area".

In the following figure, Area 2 has no direct physical link to the backbone area 0. Configuring a virtual link between ABRs can connect Area 2 to the backbone area.

Figure 75 Virtual link application 1

Another application of virtual links is to provide redundant links. If the backbone area cannot maintain internal connectivity due to a physical link failure, configuring a virtual link can guarantee logical connectivity in the backbone area, as shown below.

Figure 76 Virtual link application 2

The virtual link between the two ABRs acts as a point-to-point connection. Therefore, you can configure interface parameters such as hello packet interval on the virtual link as they are configured on physical interfaces.

The two ABRs on the virtual link exchange OSPF packets with each other directly, and the OSPF routers in between simply convey these OSPF packets as normal IP packets.

(Totally) Stub area

The ABR in a stub area does not distribute Type-5 LSAs into the area, so the routing table size and amount of routing information in this area are reduced significantly.

You can configure the stub area as a totally stub area, where the ABR advertises neither the destinations in other areas nor the external routes.

Stub area configuration is optional, and not every area is eligible to be a stub area. In general, a stub area resides on the border of the AS.

The ABR in a stub area generates a default route into the area.

Note the following when configuring a (totally) stub area:

- The backbone area cannot be a (totally) stub area.
- The **stub** command must be configured on routers in a (totally) stub area.

- A (totally) stub area cannot have an ASBR because AS external routes cannot be distributed into the stub area.
- Virtual links cannot transit (totally) stub areas.

NSSA area

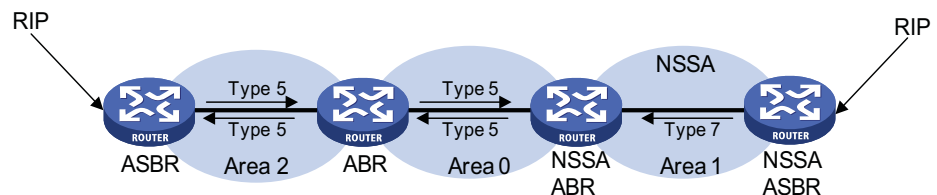
Similar to a stub area, an NSSA area imports no AS external LSA (Type-5 LSA) but can import Type-7 LSAs that are generated by the ASBR and distributed throughout the NSSA area. When traveling to the NSSA ABR, Type-7 LSAs are translated into Type-5 LSAs by the ABR for advertisement to other areas.

In the following figure, the OSPF AS contains three areas: Area 1, Area 2 and Area 0. The other two ASs employ the RIP protocol. Area 1 is an NSSA area, and the ASBR in it translates RIP routes into Type-7 LSAs and advertises them throughout Area 1. When these LSAs travel to the NSSA ABR, the ABR translates Type-7 LSAs to Type-5 LSAs for advertisement to Area 0 and Area 2.

On the left of the figure, RIP routes are translated into Type-5 LSAs by the ASBR of Area 2 and distributed into the OSPF AS. However, Area 1 is an NSSA area, so these Type-5 LSAs cannot travel to Area 1.

Like stub areas, virtual links cannot transit NSSA areas.

Figure 77 NSSA area



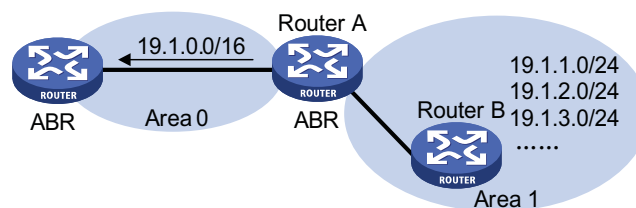
Route summarization

Route summarization: An ABR or ASBR summarizes routes with the same prefix with a single route and distribute it to other areas.

Via route summarization, routing information across areas and the size of routing tables on routers will be reduced, improving calculation speed of routers.

For example, as shown in the following figure, in Area 1 are three internal routes 19.1.1.0/24, 19.1.2.0/24, and 19.1.3.0/24. By configuring route summarization on Router A, the three routes are summarized with the route 19.1.0.0/16 that is advertised into Area 0.

Figure 78 Route summarization



OSPF has two types of route summarization:

1 ABR route summarization

To distribute routing information to other areas, an ABR generates Type-3 LSAs on a per network segment basis for an attached non-backbone area. If contiguous network segments are available in the area, you can summarize them with a single network segment. The ABR in the area distributes only the summary LSA to reduce the scale of LSDBs on routers in other areas.

2 ASBR route summarization

If summarization for redistributed routes is configured on an ASBR, it will summarize redistributed Type-5 LSAs that fall into the specified address range. If in an NSSA area, it also summarizes Type-7 LSAs that fall into the specified address range.

If this feature is configured on an ABR, the ABR will summarize Type-5 LSAs translated from Type-7 LSAs.

Route types

OSPF prioritize routes into four levels:

- Intra-area route
- Inter-area route
- Type-1 external route
- Type-2 external route

The intra-area and inter-area routes describe the network topology of the AS, while external routes describe routes to destinations outside the AS.

OSPF classifies external routes into two types: Type-1 and Type-2. A Type-1 external route is an IGP route, such as a RIP or static route, which has high credibility and whose cost is comparable with the cost of an OSPF internal route. The cost from a router to the destination of the Type-1 external route = the cost from the router to the corresponding ASBR + the cost from the ASBR to the destination of the external route.

A Type-2 external route is an EGP route, which has low credibility, so OSPF considers the cost from the ASBR to the destination of the Type-2 external route is much bigger than the cost from the ASBR to an OSPF internal router. Therefore, the cost from the internal router to the destination of the Type-2 external route = the cost from the ASBR to the destination of the Type-2 external route. If two routes to the same destination have the same cost, then take the cost from the router to the ASBR into consideration.

Classification of OSPF Networks

OSPF network types

OSPF classifies networks into four types upon the link layer protocol:

- Broadcast: When the link layer protocol is Ethernet or FDDI, OSPF considers the network type broadcast by default. On Broadcast networks, packets are sent to multicast addresses (such as 224.0.0.5 and 224.0.0.6).

- NBMA (Non-Broadcast Multi-Access): When the link layer protocol is Frame Relay, ATM or X.25, OSPF considers the network type as NBMA by default. Packets on these networks are sent to unicast addresses.
- P2MP (point-to-multipoint): By default, OSPF considers no link layer protocol as P2MP, which is a conversion from other network types such as NBMA in general. On P2MP networks, packets are sent to multicast addresses (224.0.0.5).
- P2P (point-to-point): When the link layer protocol is PPP or HDLC, OSPF considers the network type as P2P. On P2P networks, packets are sent to multicast addresses (224.0.0.5).

NBMA network configuration principle

Typical NBMA networks are ATM and Frame Relay networks.

You need to perform some special configuration on NBMA interfaces. Since these interfaces cannot broadcast hello packets for neighbor location, you need to specify neighbors manually and configure whether the neighbors have the DR election right.

An NBMA network is fully meshed, which means any two routers in the NBMA network have a direct virtual link for communication. If direct connections are not available between some routers, the type of interfaces associated should be configured as P2MP, or as P2P for interfaces with only one neighbor.

Differences between NBMA and P2MP networks:

- NBMA networks are fully meshed, non-broadcast and multi access. P2MP networks are not required to be fully meshed.
- It is required to elect the DR and BDR on NBMA networks, while DR and BDR are not available on P2MP networks.
- NBMA is the default network type, while P2MP is a conversion from other network types, such as NBMA in general.
- On NBMA networks, packets are unicast, and neighbors are configured manually on routers. On P2MP networks, packets are multicast.

DR and BDR DR/BDR introduction

On broadcast or NBMA networks, any two routers exchange routing information with each other. If n routers are present on a network, $n(n-1)/2$ adjacencies are required. Any change on a router in the network generates traffic for routing information synchronization, consuming network resources. The Designated Router is defined to solve the problem. All other routers on the network send routing information to the DR, which is responsible for advertising link state information.

If the DR fails to work, routers on the network have to elect another DR and synchronize information with the new DR. It is time-consuming and prone to routing calculation errors. The Backup Designated Router (BDR) is introduced to reduce the synchronization period.

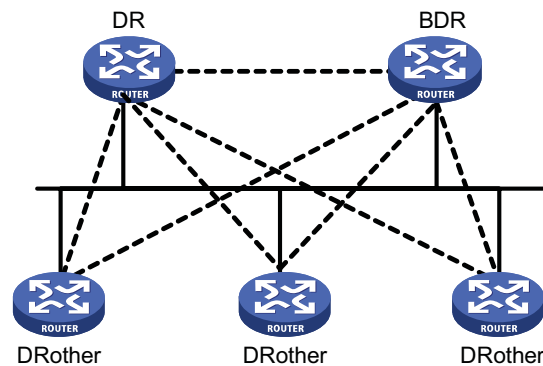
The BDR is elected along with the DR and establishes adjacencies for routing information exchange with all other routers. When the DR fails, the BDR will

become the new DR in a very short period by avoiding adjacency establishment and DR reelection. Meanwhile, other routers elect another BDR, which requires a relatively long period but has no influence on routing calculation.

Other routers, also known as DRothers, establish no adjacency and exchange no routing information with each other, thus reducing the number of adjacencies on broadcast and NBMA networks.

In the following figure, real lines are Ethernet physical links, and dashed lines represent adjacencies. With the DR and BDR in the network, only seven adjacencies are enough.

Figure 79 DR and BDR in a network



DR/BDR election

The DR and BDR in a network are elected by all routers rather than configured manually. The DR priority of an interface determines its qualification for DR/BDR election. Interfaces attached to the network and having priorities higher than '0' are election candidates.

The election votes are hello packets. Each router sends the DR elected by itself in a hello packet to all the other routers. If two routers on the network declare themselves as the DR, the router with the higher DR priority wins. If DR priorities are the same, the router with the higher router ID wins. In addition, a router with the priority 0 cannot become the DR/BDR.



- *The DR election is available on broadcast, NBMA interfaces rather than P2P, or P2MP interfaces.*
- *A DR is an interface of a router and belongs to a single network segment. The router's other interfaces may be a BDR or DRoother.*
- *After DR/BDR election and then a new router joins, it cannot become the DR immediately even if it has the highest priority on the network.*
- *The DR may not be the router with the highest priority in a network, and the BDR may not be the router with the second highest priority.*

OSPF Packet Formats

OSPF packets are directly encapsulated into IP packets. OSPF has the IP protocol number 89. The OSPF packet format is shown below (taking a LSU packet as an example).

Figure 80 OSPF packet format

IP header	OSPF packet header	Number of LSAs	LSA header	LSA Data
-----------	--------------------	----------------	------------	----------

OSPF packet header

OSPF packets are classified into five types that have the same packet header, as shown below.

Figure 81 OSPF packet header

0	7	15	31
Version	Type	Packet length	
Router ID			
Area ID			
Checksum		AuType	
Authentication			
Authentication			

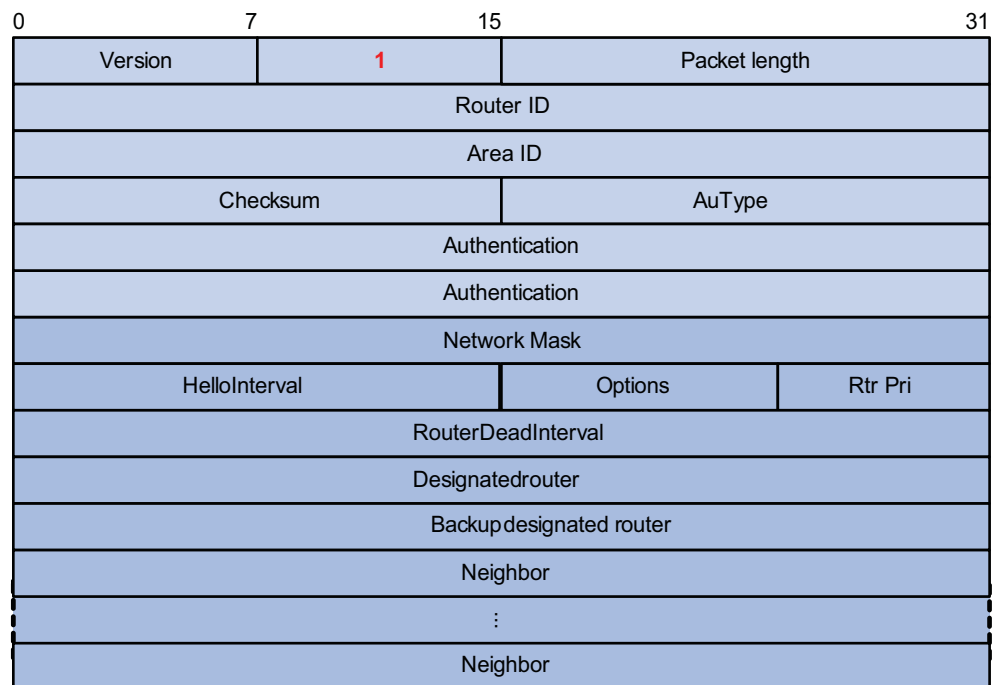
- Version: OSPF version number, which is 2 for OSPFv2.
- Type: OSPF packet type from 1 to 5, corresponding with hello, DD, LSR, LSU and LSAck respectively.
- Packet length: Total length of the OSPF packet in bytes, including the header.
- Router ID: ID of the advertising router.
- Area ID: ID of the area where the advertising router resides.
- Checksum: Checksum of the message.
- Autype: Authentication type from 0 to 2, corresponding with non-authentication, simple (plaintext) authentication and MD5 authentication respectively.
- Authentication: Information determined by authentication type. It is not defined for authentication type 0. It is defined as password information for authentication type 1, and defined as Key ID, MD5 authentication data length and sequence number for authentication type 2.



MD5 authentication data is added following an OSPF packet rather than contained in the Authentication field.

Hello packet

A router sends hello packets periodically to neighbors to find and maintain neighbor relationships and to elect the DR/BDR, including information about values of timers, DR, BDR and neighbors already known. The format is shown below:

Figure 82 Hello packet format

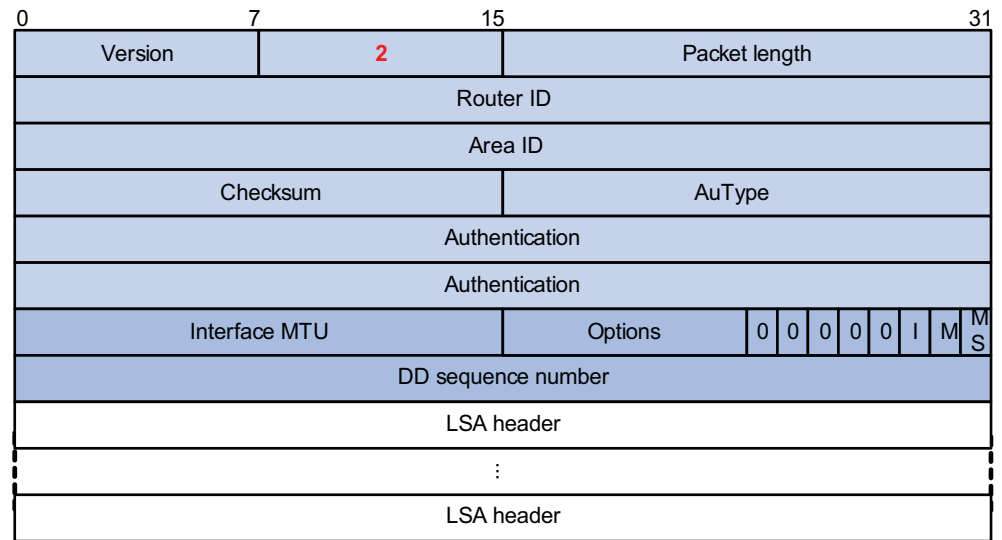
Major fields:

- Network Mask: Network mask associated with the router's sending interface. If two routers have different network masks, they cannot become neighbors.
- HelloInterval: Interval for sending hello packets. If two routers have different intervals, they cannot become neighbors.
- Rtr Pri: Router priority. A value of 0 means the router cannot become the DR/BDR.
- RouterDeadInterval: Time before declaring a silent router down. If two routers have different time values, they cannot become neighbors.
- Designated Router: IP address of the DR interface.
- Backup Designated Router: IP address of the BDR interface
- Neighbor: Router ID of the neighbor router.

DD packet

Two routers exchange database description (DD) packets describing their LSDBs for database synchronization, contents in DD packets including the header of each LSA (uniquely representing a LSA). The LSA header occupies small part of an LSA to reduce traffic between routers. The recipient checks whether the LSA is available using the LSA header.

The DD packet format:

Figure 83 DD packet format

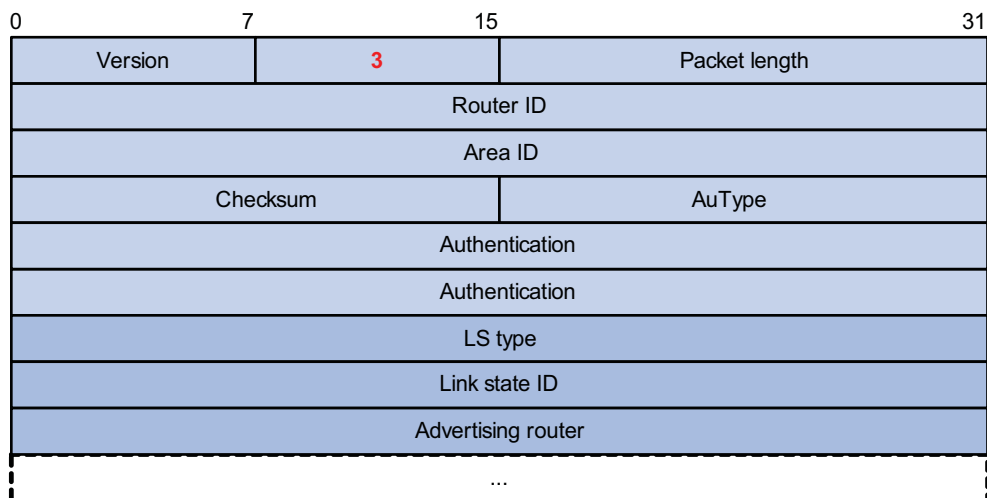
Major fields:

- Interface MTU: Size in bytes of the largest IP datagram that can be sent out the associated interface, without fragmentation.
- I (Initial) The Init bit, which is set to 1 if the packet is the first packet of database description packets, and set to 0 if not.
- M (More): The More bit, which is set to 0 if the packet is the last packet of DD packets, and set to 1 if more DD Packets are to follow.
- MS (Master/Slave): The Master/Slave bit. When set to 1, it indicates that the router is the master during the database exchange process. Otherwise, the router is the slave.
- DD Sequence Number: Used to sequence the collection of database description packets for ensuring reliability and intactness of DD packets between the master and slave. The initial value is set by the master. The DD sequence number then increments until the complete database description has been sent.

LSR packet

After exchanging DD packets, any two routers know which LSAs of the peer routers are missing from the local LSDBs. In this case, they send LSR (link state request) packets, requesting the missing LSAs. The packets contain the digests of the missing LSAs. The following figure shows the LSR packet format.

Figure 84 LSR packet format



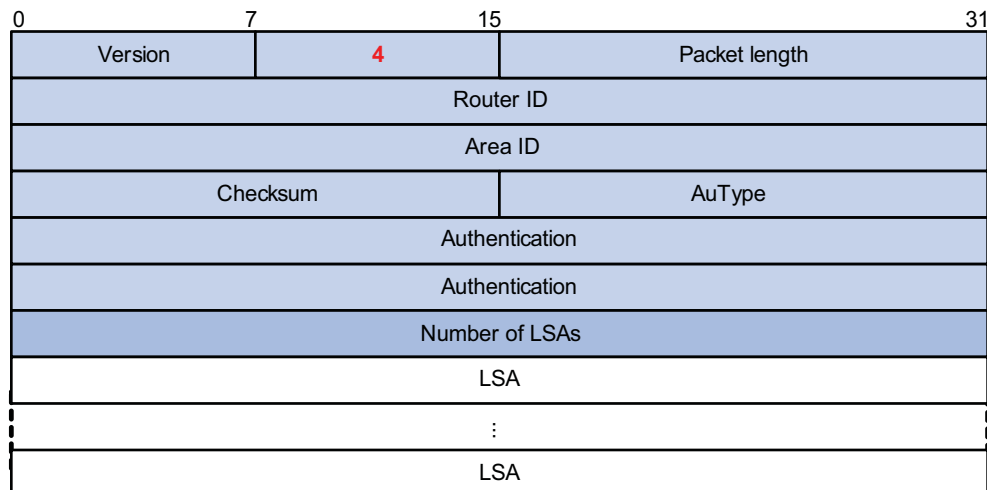
Major fields:

- LS type: Type number of the LSA to be requested. Type 1 for example indicates the Router LSA.
- Link State ID: Determined by LSA type.
- Advertising Router: ID of the router that sent the LSA.

LSU packet

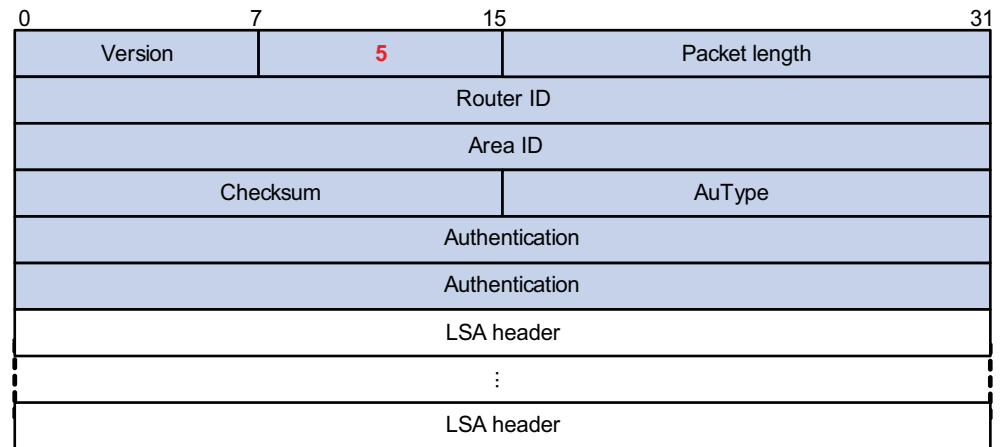
LSU (Link State Update) packets are used to send the requested LSAs to peers, and each packet carries a collection of LSAs. The LSU packet format is shown below.

Figure 85 LSU packet format

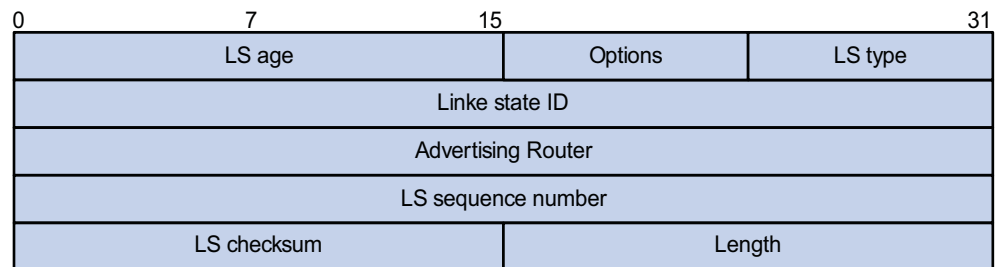


LSAck packet

LSAck (Link State Acknowledgment) packets are used to acknowledge received LSU packets, contents including LSA headers to describe the corresponding LSAs. Multiple LSAs can be acknowledged in a single Link State Acknowledgment packet. The following figure gives its format.

Figure 86 LSAck packet format**LSA header format**

All LSAs have the same header, as shown in the following figure.

Figure 87 LSA header format

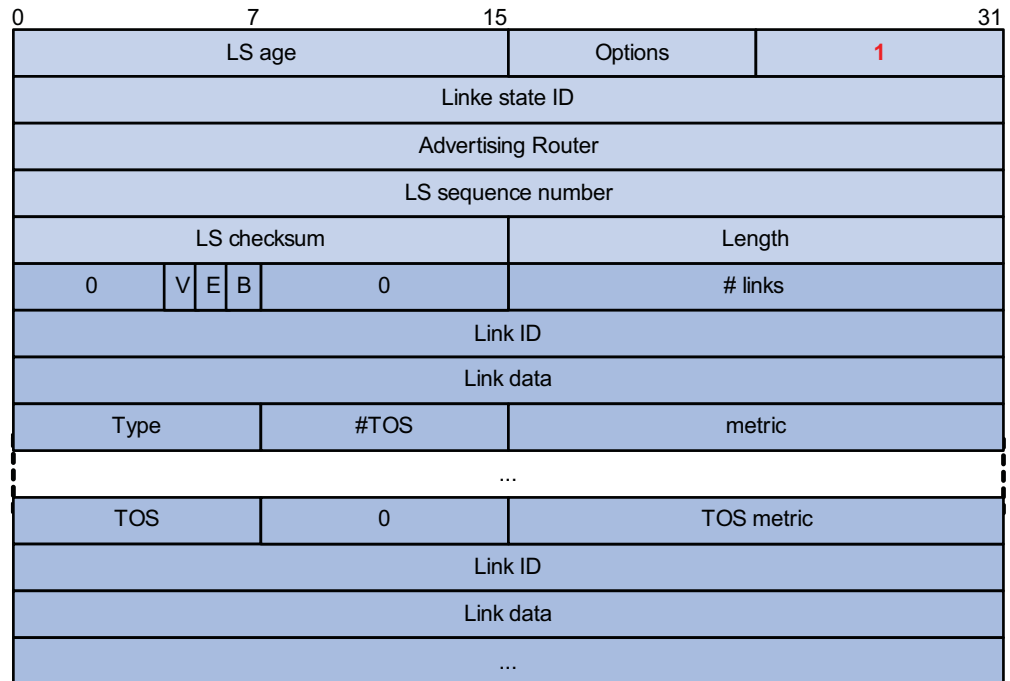
Major fields:

- LS age: Time in seconds elapsed since the LSA was originated. A LSA ages in the LSDB (added by 1 per second), but does not in transmission.
- LS type: Type of the LSA.
- Link State ID: The contents of this field depend on the LSA's type
- LS sequence number: Used by other routers to judge new and old LSAs.
- LS checksum: Checksum of the LSA except the LS age field.
- Length: Length in bytes of the LSA, including the LSA header.

Formats of LSAs

1 Router LSA

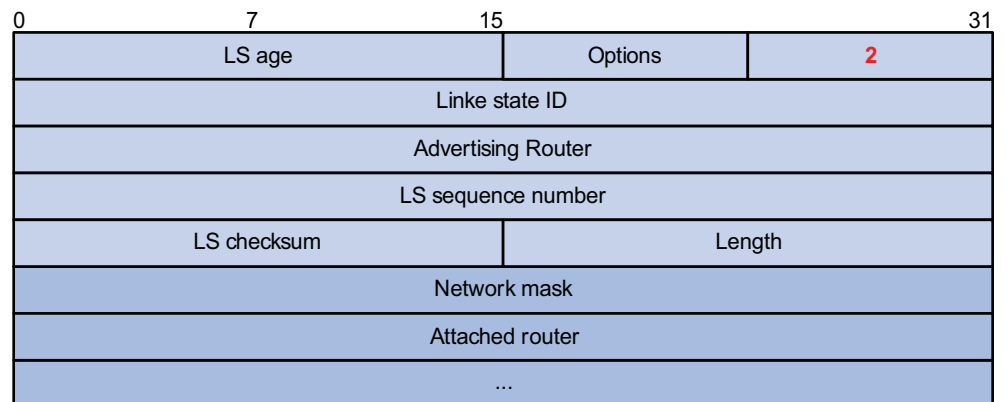
Figure 88 Router LSA format



Major fields:

- Link State ID: ID of the router that originated the LSA.
- V (Virtual Link): Set to 1 if the router that originated the LSA is a virtual link endpoint.
- E (External): Set to 1 if the router that originated the LSA is an ASBR.
- B (Border): Set to 1 if the router that originated the LSA is an ABR.
- # links: Number of router links (interfaces) to the area, described in the LSA.
- Link ID: Determined by Link type.
- Link Data: Determined by Link type.
- Type: Link type. A value of 1 indicates a point-to-point link to a remote router; a value of 2 indicates a link to a transit network; a value of 3 indicates a link to a stub network; a value of 4 indicates a virtual link.
- #TOS: Number of different TOS metrics given for this link.
- metric: Cost of using this router link.
- TOS: IP Type of Service that this metric refers to.
- TOS metric: TOS-specific metric information.
- Network LSA

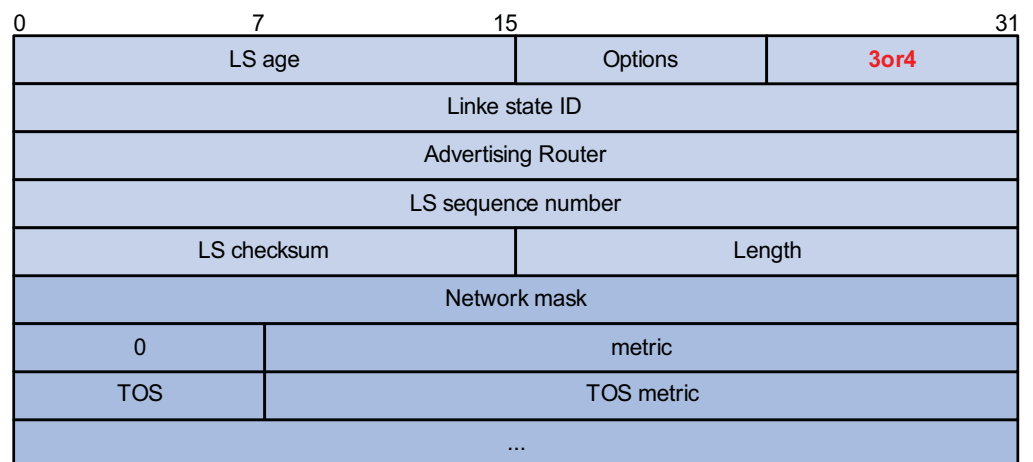
A Network LSA is originated by the DR on a broadcast or NBMA network. The LSA describes all routers attached to the network.

Figure 89 Network LSA format

Major fields:

- Link State ID: The interface address of the DR
- Network Mask: The mask of the network (a broadcast or NBMA network)
- Attached Router: The IDs of the routers, which are adjacent to the DR, including the DR itself
- Summary LSA

Network summary LSAs (Type-3 LSAs) and ASBR summary LSAs (Type-4 LSAs) are originated by ABRs. Other than the difference in the Link State ID field, the format of type 3 and 4 summary-LSAs is identical.

Figure 90 Summary LSA format

Major fields:

- Link State ID: For a Type-3 LSA, it is an IP address outside the area; for a type 4 LSA, it is the router ID of an ASBR outside the area.
- Network Mask: The network mask for the type 3 LSA; set to 0.0.0.0 for the Type-4 LSA
- metric: The metric to the destination

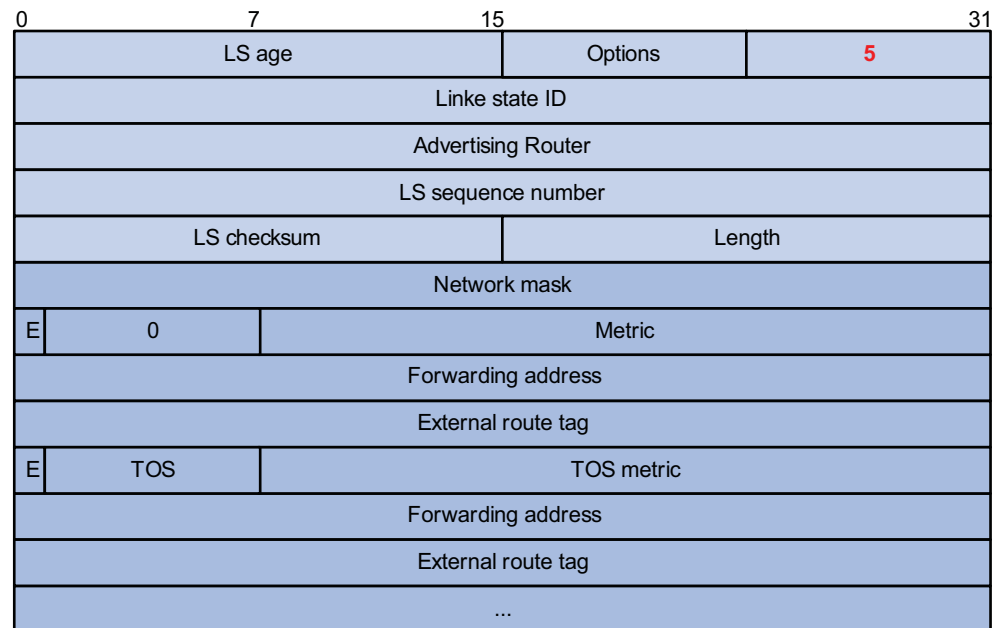


A Type-3 LSA can be used to advertise a default route, having the Link State ID and Network Mask set to 0.0.0.0.

1 AS external LSA

An AS external LSA originates from an ASBR, describing routing information to a destination outside the AS.

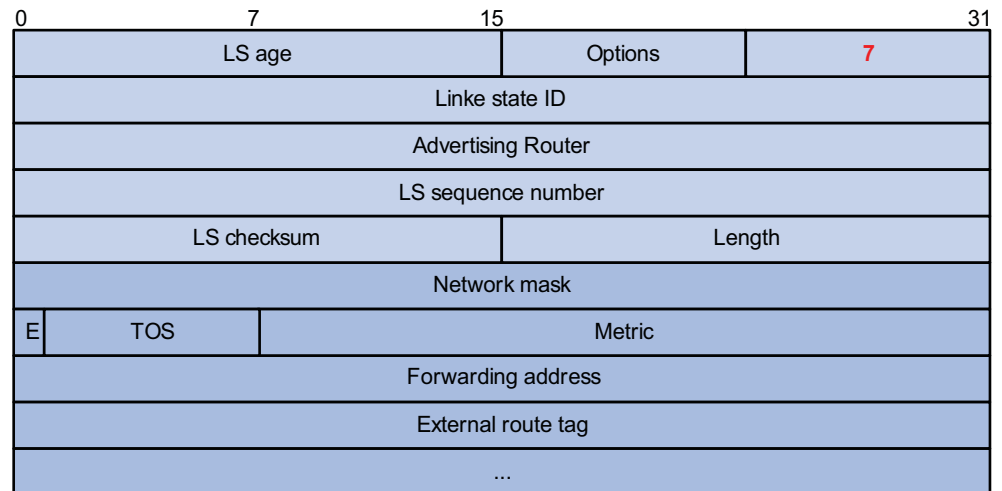
Figure 91 AS external LSA format



Major fields:

- Link State ID: The IP address of another AS to be advertised. When describing a default route, the Link State ID is always set to Default Destination (0.0.0.0) and the Network Mask is set to 0.0.0.0
- Network Mask: The IP address mask for the advertised destination
- E (External Metric): The type of the external metric value, which is set to 1 for type 2 external routes, and set to 0 for type 1 external routes. Refer to “Route types” on page 280 for description about external route types
- metric: The metric to the destination
- Forwarding Address: Data traffic for the advertised destination will be forwarded to this address
- External Route Tag: A tag attached to each external route. This is not used by the OSPF protocol itself. It may be used to manage external routes.
- NSSA external LSA

An NSSA external LSA originates from the ASBR in a NSSA and is flooded in the NSSA area only. It has the same format as the AS external LSA.

Figure 92 NSSA external LSA format

Supported OSPF Features

Multi-process

With multi-process support, multiple OSPF processes can run on a router simultaneously and independently. Routing information interactions between different processes seem like interactions between different routing protocols. Multiple OSPF processes can use the same RID.

An interface of a router can only belong to a single OSPF process.

Authentication

OSPF supports authentication on packets. Only packets that pass the authentication are received. If hello packets cannot pass authentication, no neighbor relationship can be established.

The authentication type for interfaces attached to a single area must be identical. Authentication types include non-authentication, plaintext authentication and MD5 ciphertext authentication. The authentication password for interfaces attached to a network segment must be identical.

OSPF Graceful Restart



For GR information, refer to "GR Overview" on page 247.

After an OSPF GR Restarter restarts OSPF, it needs to perform the following two tasks in order to re-synchronize its LSDB with its neighbors.

- To obtain once again effective OSPF neighbor information, supposing the adjacencies are not changed.
- To obtain once again LSDB contents.

Before the restart, the GR Restarter originates Grace-LSAs to negotiate the GR capability. During the restart, the GR Helpers continue to advertise their adjacencies with the GR Restarter.

After the restart, the GR Restarter will send an OSPF GR signal to its neighbors that will not reset their adjacencies with it. In this way, the GR Restarter can restore the neighbor table upon receiving the responses from neighbors.

After reestablishing neighbor relationships, the GR Restarter will synchronize the LSDB and exchange routing information with all adjacent GR-capable neighbors. After that, the GR Restarter will update its own routing table and forwarding table based on the new routing information and remove the stale routes. In this way, the OSPF routing convergence is complete.

Protocols and Standards

- RFC 1765: OSPF Database Overflow
- RFC 2328: OSPF Version 2
- RFC 3101: OSPF Not-So-Stubby Area (NSSA) Option
- RFC 3137: OSPF Stub Router Advertisement

OSPF Configuration Task List

Complete the following tasks to configure OSPF:

Task	Remarks
"Configuring OSPF Basic Functions" on page 293	Required
"Configuring OSPF Area Parameters" on page 294	Optional
"Configuring OSPF Network Types" on page 295	Optional
	"Configuring the OSPF Network Type for an Interface" on page 296
	"Configuring an NBMA Neighbor" on page 296
	"Configuring a Router Priority for an OSPF Interface" on page 296
"Configuring OSPF Route Control" on page 297	Optional
	"Configuring OSPF Route Summarization" on page 297
	"Configuring OSPF Inbound Route Filtering" on page 298
	"Configuring ABR Type-3 LSA Filtering" on page 298
	"Configuring an OSPF Cost for an Interface" on page 298
	"Configuring the Maximum Number of OSPF Routes" on page 299
	"Configuring the Maximum Number of Load-balanced Routes" on page 299
	"Configuring a Priority for OSPF" on page 299
	"Configuring OSPF Route Redistribution" on page 300

Task	Remarks	
"Configuring OSPF Network Optimization" on page 300	"Configuring OSPF Packet Timers" on page 301	Optional
	"Specifying an LSA Transmission Delay" on page 302	Optional
	"Specifying SPF Calculation Interval" on page 302	Optional
	"Specifying the LSA Minimum Repeat Arrival Interval" on page 302	Optional
	"Specifying the LSA Generation Interval" on page 303	Optional
	"Disabling Interfaces from Sending OSPF Packets" on page 303	Optional
	"Configuring Stub Routers" on page 304	Optional
	"Configuring OSPF Authentication" on page 304	Optional
	"Adding the Interface MTU into DD Packets" on page 305	Optional
	"Configuring the Maximum Number of External LSAs in LSDB" on page 305	Optional
	"Making External Route Selection Rules Defined in RFC 1583 Compatible" on page 305	Optional
	"Logging Neighbor State Changes" on page 305	Optional
	"Configuring OSPF Network Management" on page 305	Optional
	"Enabling the Advertisement and Reception of Opaque LSAs" on page 306	Optional
"Configuring OSPF Graceful Restart" on page 306	"Configuring the GR Capability" on page 306	Optional
	"Configuring the OSPF GR Helper" on page 307	Optional
	"Triggering OSPF Graceful Restart" on page 308	Optional

Configuring OSPF Basic Functions

You need to enable OSPF, specify an interface and area ID first before performing other tasks.

Prerequisites

Before configuring OSPF, you need to configure IP addresses for interfaces, making neighboring nodes accessible with each other at the network layer.

Configuration Procedure

- Configure a Router ID

To ensure OSPF stability, you need to decide on router IDs and configure them manually. Any two routers in an AS must have different IDs. In practice, the ID of a router is the IP address of one of its interfaces.

- Enable an OSPF process

The system supports OSPF multi-process. When a router runs multiple OSPF processes, you need to specify an ID for each process, which takes effect locally and has no influence on packet exchange between routers. Therefore, two routers having different process IDs can exchange packets.

- Configure an area and specify networks in the area

The configurations for routers in an area are performed on the area basis. Wrong configurations may cause communication failures, even routing information block or routing loops between neighboring routers.

Follow these steps to configure OSPF basic functions:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable OSPF and enter its view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	Required
Configure a description for the OSPF process	description <i>description</i>	Optional Not configured by default
Configure an OSPF area and enter OSPF area view	area <i>area-id</i>	Required Not configured by default
Configure a description for the area	description <i>description</i>	Optional Not configured by default
Specify a network to enable OSPF on the interface attached to the network	network <i>ip-address</i> <i>wildcard-mask</i>	Required Not configured by default



- *An OSPF process ID is unique.*
- *A network segment can only belong to one area.*
- *It is recommended to configure a description for each OSPF process to help identify purposes of processes and for ease of management and memorization.*
- *It is recommended to configure a description for each area to help identify purposes of areas and for ease of management and memorization.*

Configuring OSPF Area Parameters

Splitting an OSPF AS into multiple areas reduces the number of LSAs in the networks and extends the OSPF application. For those non-backbone areas residing on the AS boundary, you can configure them as stub areas to further reduce the size of routing tables on routers in these areas and the number of LSAs.

A stub area cannot redistribute routes, and for this reason, NSSA was introduced. In NSSA areas, Type-7 LSAs (NSSA External LSAs) can be advertised. Type 7 LSAs originate from the ASBR in a NSSA area. When arriving at the ABR in the NSSA

area, these LSAs will be translated into type 5 LSAs for advertisement to other areas.

Non-backbone areas exchange routing information via the backbone area. Therefore, the backbone and non-backbone areas, including the backbone itself must maintain connectivity.

If necessary physical links are not available for this connectivity maintenance, you can configure virtual links to solve it.

Prerequisites Before configuring an OSPF area, you have configured:

- IP addresses for interfaces, making neighboring nodes accessible with each other at the network layer.
- OSPF basic functions.

Configuration Procedure Follow these steps to configure OSPF area parameters:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Enter area view	area <i>area-id</i>	-
Configure the area as a stub area	stub [no-summary]	Optional Not configured by default
Configure the area as an NSSA area	nssa [default-route-advertise no-import-route no-summary] *	Optional Not configured by default
Specify a cost for the default route advertised to the stub or NSSA area	default-cost <i>cost</i>	Optional Defaults to 1.
Configure a virtual link	vlink-peer <i>router-id</i> [hello <i>seconds</i> retransmit <i>seconds</i> trans-delay <i>seconds</i> dead <i>seconds</i> simple [plain cipher] <i>password</i>] { md5 hmac-md5 } <i>key-id</i> [plain cipher] <i>password</i>] *	Optional Configured on both ends of a virtual link Note that hello and dead parameters must be identical on both ends of the link.
Advertise a host route	host-advertise <i>ip-address</i> <i>cost</i>	Optional Not advertised by default



- It is required to use the **stub** command on routers attached to a stub area.
- It is required to use the **nssa** command on routers attached to an NSSA area.
- Using the **default-cost** command only takes effect on the ABR of a stub area or the ABRIASBR of an NSSA area.

Configuring OSPF Network Types

OSPF classifies networks into four types upon link layer protocols. Since an NBMA network must be fully meshed, namely, any two routers in the network must have a virtual link in between. In most cases, however, the requirement cannot be satisfied, so you need to change the network type using commands.

For routers having no direct link in between, you can configure the P2MP type for the related interfaces. If a router in the NBMA network has only a single peer, you can configure the P2P type for the related interfaces.

In addition, when configuring broadcast and NBMA networks, you can specify for interfaces router priorities for DR/BDR election. In practice, the routers having higher reliability should become the DR/BDR.

Prerequisites Before configuring OSPF network types, you have configured:

- IP addresses for interfaces, making neighboring nodes accessible with each other at network layer.
- OSPF basic functions.

Configuring the OSPF Network Type for an Interface

Follow these steps to configure the OSPF network type for an interface:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure a network type	ospf network-type { broadcast nbma p2mp p2p }	Optional Not configured by default The network type of an interface depends on the media type of the interface.



- *Configuring a new network type for an interface overwrites the previous one (if any).*
- *If the two interfaces on a link are both configured as the broadcast, NBMA or P2MP type, they can not establish the neighbor relationship unless they are on the same network segment.*

Configuring an NBMA Neighbor

For NBMA interfaces that cannot broadcast hello packets to find neighbors, you need to specify the IP addresses and DR priorities of neighbors manually.

Follow these steps to configure a neighbor and its DR priority:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Specify an NBMA neighbor and its DR priority	peer <i>ip-address</i> [dr-priority <i>dr-priority</i>]	Required

Configuring a Router Priority for an OSPF Interface

For broadcast or NBMA interfaces, you can configure router priorities for DR/BDR election.

Follow these steps to configure a router priority for an OSPF interface:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure a router priority for the interface	ospf dr-priority <i>priority</i>	Optional The default router priority is 1.



The DR priority configured with the **ospf dr-priority** command and the one with the **peer** command have the following differences

- The former is for actual DR election.
- The latter is to indicate whether a neighbor has the election right or not. If you configure the DR priority for a neighbor as 0, the local router will consider the neighbor has no election right, and thus no hello packet is sent to this neighbor, reducing the number of hello packets for DR/BDR election on networks. However, if the local router is the DR or BDR, it sends hello packets to the neighbor with priority 0 for adjacency establishment.

Configuring OSPF Route Control

This section covers how to control OSPF routing information advertisement and reception, and route redistribution from other protocols.

Prerequisites

Before configuring this task, you have configured:

- IP addresses for interfaces
- OSPF basic functions
- Corresponding filters if routing information filtering is needed.

Configuring OSPF Route Summarization

OSPF route summarization includes:

- Configuring route summarization between OSPF areas on an ABR
- Configuring route summarization when redistributing routes into OSPF on an ASBR

Follow these steps to configure route summarization between OSPF areas on an ABR:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Enter OSPF area view	area <i>area-id</i>	Required
Configure ABR route summarization	abr-summary <i>ip-address</i> { <i>mask</i> <i>mask-length</i> } [advertise not-advertise] [cost <i>cost</i>]	Required Available on an ABR only Not configured by default

Follow these steps to configure route summarization when redistributing routes into OSPF on an ASBR:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Configure ASBR route summarization	asbr-summary <i>ip-address</i> { <i>mask</i> <i>mask-length</i> } [tag <i>tag</i> not-advertise cost <i>cost</i>] *	Required Available on an ASBR only Not configured by default

Configuring OSPF Inbound Route Filtering

Follow these steps to configure inbound route filtering:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	Required
Configure inbound route filtering	filter-policy { <i>acl-number</i> ip-prefix <i>ip-prefix-name</i> gateway <i>ip-prefix-name</i> } import	Required Not configured by default



Since OSPF is a link state-based interior gateway protocol, routing information is contained in LSAs. However, OSPF cannot filter LSAs. Using the **filter-policy import** command is to filter routes computed by OSPF, and only routes not filtered out are installed into the routing table.

Configuring ABR Type-3 LSA Filtering

Follow these steps to configure Type-3 LSA filtering on an ABR:

To do...	Use the command...	Remarks
Enter system view	System-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Enter area view	area <i>area-id</i>	-
Configure ABR Type-3 LSA filtering	filter { <i>acl-number</i> ip-prefix <i>ip-prefix-name</i> } { import export }	Required Not configured by default

Configuring an OSPF Cost for an Interface

Follow these steps to configure an OSPF cost for an interface:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure an OSPF cost for the interface	ospf cost <i>value</i>	Optional By default, an interface computes its cost according to the bandwidth. The cost value defaults to 1 for VLAN interfaces.

Follow these steps to configure a bandwidth reference value:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Configure a bandwidth reference value	bandwidth-reference <i>value</i>	Optional The value defaults to 100 Mbps.



If no OSPF cost is configured for an interface, OSPF computes the cost automatically: $\text{Interface OSPF cost} = \text{Bandwidth reference value} / \text{Interface bandwidth}$. If the calculated cost is greater than 65535, the value of 65535 is used.

Configuring the Maximum Number of OSPF Routes

Follow these steps to configure the maximum number of routes:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Configure the maximum number of OSPF routes	maximum-routes { external inter intra } <i>number</i>	Optional The default number is 12288.

Configuring the Maximum Number of Load-balanced Routes

If several routes with the same cost to the same destination are available, configuring them as load-balanced routes can improve link utilization.

Follow these steps to configure the maximum number of load-balanced routes:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Configure the maximum number of equivalent load-balanced routes	maximum load-balancing <i>maximum</i>	Optional The default number is 4.

Configuring a Priority for OSPF

A router may run multiple routing protocols, and it sets a priority for each protocol. When a route found by several routing protocols, the route found by the protocol with the highest priority will be selected.

Follow these steps to configure a priority for OSPF:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-

To do...	Use the command...	Remarks
Configure a priority for OSPF	preference [ase] [route-policy <i>route-policy-name</i>] <i>value</i>	Optional The priority of OSPF internal routes defaults to 10. The priority of OSPF external routes defaults to 150.

Configuring OSPF Route Redistribution

Follow these steps to configure OSPF route redistribution:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Configure OSPF to redistribute routes from another protocol	import-route <i>protocol</i> [<i>process-id</i> allow-ibgp] [cost <i>cost</i> type <i>type</i> tag <i>tag</i> route-policy <i>route-policy-name</i>]*	Required Not configured by default
Configure OSPF to filter redistributed routes before advertisement	filter-policy { <i>acl-number</i> ip-prefix <i>ip-prefix-name</i> } export [<i>protocol</i> [<i>process-id</i>]]	Optional Not configured by default
Redistribute a default route	default-route-advertise [always cost <i>cost</i> type <i>type</i> route-policy <i>route-policy-name</i>]* default-route-advertise summary <i>cost</i> <i>cost</i>	Optional Not redistributed by default
Configure the default parameters for redistributed routes (cost, route number, tag and type)	default { cost <i>cost</i> limit <i>limit</i> tag <i>tag</i> type <i>type</i> } *	Optional By default, the default cost is 1, default upper limit of routes redistributed per time is 1000, default tag is 1 and default type of redistributed routes is Type-2.



- Using the **import-route** command cannot redistribute a default external route. To do so, you need to use the **default-route-advertise** command.
- The **default-route-advertise summary cost** command is applicable only to VPN, and the default route is redistributed in a Type-3 LSA. The PE router will advertise the default route to the CE router. The switch does not support this command currently because the switch does not support VPN.
- By filtering redistributed routes, OSPF adds only routes, which are not filtered out, into Type-5 LSAs or Type-7 LSAs for advertisement.
- You can configure default parameters such as the cost, upper limit, tag and type for redistributed routes. Tags are used to identify information related to protocols. For example, when redistributing BGP routes, OSPF uses AS IDs as route tags.

Configuring OSPF Network Optimization

You can optimize your OSPF network in the following ways:

- Change OSPF packet timers to adjust the OSPF network convergence speed and network load. On low speed links, you need to consider the delay time for sending LSAs on interfaces.
- Change the interval for SPF calculation to reduce resource consumption caused by frequent network changes.
- Configure OSPF authentication to meet high security requirements of some mission-critical networks.
- Configure OSPF network management functions, such as binding OSPF MIB with a process, sending trap information and collecting log information.

Prerequisites Before configuring OSPF network optimization, you have configured:

- IP addresses for interfaces;
- OSPF basic functions.

Configuring OSPF Packet Timers

You can configure the following timers on OSPF interfaces as needed:

- Hello timer: Interval for sending hello packets. It must be identical on OSPF neighbors. The longer the interval, the lower convergence speed and smaller network load.
- Poll timer: Interval for sending hello packets to the neighbor that is down on the NBMA network.
- Dead timer: Interval within which if the interface receives no hello packet from the neighbor, it declares the neighbor is down.
- LSA retransmission timer: Interval within which if the interface receives no acknowledgement packets after sending a LSA to the neighbor, it will retransmit the LSA.

Follow these steps to configure timers for OSPF packets:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Specify the hello interval	ospf timer hello <i>seconds</i>	Optional The hello interval on P2P, Broadcast interfaces defaults to 10 seconds and defaults to 30 seconds on P2MP and NBMA interfaces.
Specify the poll interval	ospf timer poll <i>seconds</i>	Optional The poll interval defaults to 120 seconds.
Specify the dead interval	ospf timer dead <i>seconds</i>	Optional The default dead interval is 40 seconds on P2P, Broadcast interfaces and 120 seconds on P2MP and NBMA interfaces.
Specify the retransmission interval	ospf timer retransmit <i>interval</i>	Optional The retransmission interval defaults to 5 seconds.



- The hello and dead intervals restore to default values after you change the network type for an interface.
- The dead interval should be at least four times the hello interval on an interface.
- The poll interval is at least four times the hello interval.
- The retransmission interval should not be so small for avoidance of unnecessary LSA retransmissions. In general, this value is bigger than the round-trip time of a packet between two adjacencies.

Specifying an LSA Transmission Delay

Since OSPF packets need time for traveling on links, extending LSA age time with a delay is necessary, especially for low speed links.

Follow these steps to specify an LSA transmission delay on an interface:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type interface-number</i>	-
Specify an LSA transmission delay	ospf trans-delay <i>seconds</i>	Optional 1 second by default

Specifying SPF Calculation Interval

The LSDB changes lead to SPF calculations. When an OSPF network changes frequently, a large amount of network resources will be occupied, reducing the working efficiency of routers. You can adjust the SPF calculation interval for the network to reduce negative influence.

Follow these steps to configure SPF calculation interval:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Specify SPF calculation interval(s)	spf-schedule-interval <i>maximum-interval</i> [<i>minimum-interval</i> [<i>incremental-interval</i>]]	Optional By default, the interval is 5 seconds.



With this task configured, when network changes are not frequent, SPF calculation applies at the minimum-interval. If network changes become frequent, SPF calculation interval is incremented by $incremental-interval \times 2^{n-2}$ (n is the number of calculation times) each time a calculation occurs, up to the maximum-interval.

Specifying the LSA Minimum Repeat Arrival Interval

After receiving the same LSA as the previously received LSA within the LSA minimum repeat arrival interval, an interface discards the LSA.

Follow these steps to configure the LSA minimum repeat arrival interval:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Configure the LSA minimum repeat arrival interval	lsa-arrival-interval <i>interval</i>	Optional Defaults to 1000 milliseconds.



The interval set with the **lsa-arrival-interval** command should be smaller or equal to the interval set with the **lsa-generation-interval** command.

Specifying the LSA Generation Interval

With this feature configured, you can protect network resources and routers from being over consumed due to frequent network changes.

Follow these steps to configure the LSA generation interval:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	Required
Configure the LSA generation interval	lsa-generation-interval <i>maximum-interval</i> [<i>initial-interval</i> [<i>incremental-interval</i>]]	Optional By default, the maximum interval is 5 seconds, the minimum interval is 0 milliseconds and the incremental interval is 5000 milliseconds.



With this command configured, when network changes are not frequent, LSAs are generated at the *minimum-interval*. If network changes become frequent, LSA generation interval is incremented by $incremental-interval \times 2^{n-2}$ (*n* is the number of generation times) each time a generation occurs, up to the *maximum-interval*.

Disabling Interfaces from Sending OSPF Packets

Follow these steps to disable interfaces from sending routing information:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Disable interfaces from sending OSPF packets	silent-interface { all <i>interface-type</i> <i>interface-number</i> }	Optional Not disabled by default



- Different OSPF processes can disable the same interface from sending OSPF packets. Use of the **silent-interface** command disables only the interfaces associated with the current process rather than interfaces associated with other processes.
- After an OSPF interface is set to silent, other interfaces on the router can still advertise direct routes of the interface in Router LSAs, but no OSPF packet can be advertised for the interface to find a neighbor. This configuration can enhance adaptability of OSPF networking and reduce resource consumption.

Configuring Stub Routers

A stub router is used for traffic control. It tells other OSPF routers not to use it to forward data, but they can have a route to it.

The Router LSAs from the stub router may contain different link type values. A value of 3 means a link to the stub network, so the cost of the link remains unchanged. A value of 1, 2 or 4 means a point-to-point link, a link to a transit network or a virtual link. In such cases, a maximum cost value of 65535 is used. Thus, other neighbors find the links to the stub router have such big costs, they will not send packets to the stub router for forwarding as long as there is a route with a smaller cost.

Follow these steps to configure a router as a stub router:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Configure the router as a stub router	stub-router	Required Not configured by default



A stub router has nothing to do with a stub area.

Configuring OSPF Authentication

By supporting packet authentication, OSPF receives packets that pass the authentication only, so failed packets cannot establish neighboring relationships.

Follow these steps to configure OSPF authentication:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Enter area view	area <i>area-id</i>	-
Configure the authentication mode	authentication-mode { simple md5 }	Required Not configured by default
Exit to OSPF view	quit	-
Exit to system view	quit	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure the authentication mode (simple authentication) for the interface	ospf authentication-mode simple [plain cipher] <i>password</i>	Optional Not configured by default
Configure the authentication mode (MD5 authentication) for the interface	ospf authentication-mode { md5 hmac-md5 } <i>key-id</i> [plain cipher] <i>password</i>	



The authentication mode and password for all interfaces attached to the same area must be identical.

Adding the Interface MTU into DD Packets

Generally, when an interface sends a DD packet, it adds 0 into the Interface MTU field of the DD packet rather than the interface MTU.

Follow these steps to add the interface MTU into DD packets:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable OSPF to add the interface MTU into DD packets	ospf mtu-enable	Optional Not enabled by default; that is, the interface fills in a value of 0.

Configuring the Maximum Number of External LSAs in LSDB

Follow these steps to configure the maximum number of external LSAs in the Link State Database:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Specify the maximum number of external LSAs in the LSDB	lsdb-overflow-limit <i>number</i>	Optional No limitation by default

Making External Route Selection Rules Defined in RFC1583 Compatible

The selection of an external route from multiple LSAs defined in RFC2328 is different from the one defined in RFC1583.

Follow these steps to make them compatible:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	Required
Make RFC1583 compatible	rfc1583 compatible	Optional Compatible by default

Logging Neighbor State Changes

Follow these steps to enable the logging of neighbor state changes:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Enable the logging of neighbor state changes	log-peer-change	Optional Enabled by default

Configuring OSPF Network Management

Follow these steps to configure OSPF network management:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Bind OSPF MIB to an OSPF process	ospf mib-binding <i>process-id</i>	Optional The first OSPF process is bound with OSPF MIB by default.
Enable OSPF trap	snmp-agent trap enable ospf [<i>process-id</i>] [ifauthfail ifcfgerror ifrxbadpkt ifstatechange iftxretransmit lsdbapproachoverflow lsdboverflow maxagelsa nbrstatechange originatelsa vifcfgerror virifauthfail virifrxbadpkt virifstatechange viriftxretransmit virnbrstatechange] *	Optional Enabled by default
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Enable messages logging	enable log [config error state]	Optional Not enabled by default

Enabling the Advertisement and Reception of Opaque LSAs

With this feature enabled, the OSPF router can receive and advertise Type 9, Type 10 and Type 11 opaque LSAs.

Follow these steps to enable the advertisement and reception of opaque LSAs:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Enable the advertisement and reception of opaque LSAs	opaque-capability enable	Optional Disabled by default

Configuring OSPF Graceful Restart

Configuring the GR Capability

You can configure the IETF standard or non IETF standard OSPF Graceful Restart capability.

Configuring the IETF standard OSPF GR capability

Follow these steps to configure the standard IETF OSPF GR capability:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPF view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Enable the advertisement and reception of opaque LSAs	opaque-capability enable	Required Disabled by default

To do...	Use the command...	Remarks
Enable the IETF standard Graceful Restart capability for OSPF	graceful-restart ietf	Optional Disabled by default
Configure the Graceful Restart interval for OSPF	graceful-restart interval <i>timer</i>	Optional 120 seconds by default



- With the **graceful-restart ietf** command used, a device can act as a GR Restarter and a GR Helper.
- Without the **graceful-restart ietf** command used, a device can only act as a GR Helper.

Configure the non-IETF standard OSPF GR capability

Follow these steps to configure non-IETF standard OSPF GR capability:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable OSPF and enter its view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	-
Enable the use of link-local signaling	enable link-local-signaling	Required Disabled by default
Enable out-of-band re-synchronization	enable out-of-band-resynchronization	Required Disabled by default
Enable non IETF standard Graceful Restart capability for OSPF	graceful-restart [nonstandard]	Optional Disabled by default
Configure Graceful Restart interval for OSPF	graceful-restart interval <i>timer</i>	Optional 120 seconds by default



- With the **graceful-restart** command used, a device can act as a GR Restarter and a GR Helper.
- Without the **graceful-restart** command used, a device can only act as a GR Helper.

Configuring the OSPF GR Helper

Follow these steps to configure the OSPF GR Helper:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable OSPF and enter its view	ospf [<i>process-id</i> router-id <i>router-id</i>] *	Required Disabled by default
Enable OSPF local link signaling	enable link-local-signaling	Required Disabled by default
Enable OSPF out of band synchronization	enable out-of-band-resynchronization	Required Disabled by default

To do...	Use the command...	Remarks
Configure for which OSPF neighbors the current router can serve as a GR Helper	graceful-restart help { <i>acl-number</i> prefix <i>prefix-list</i> }	Optional The router can server as a GR Helper for any OSPF neighbor by default.

Triggering OSPF Graceful Restart

Performing the following configuration on an OSPF router will trigger OSPF Graceful Restart. Ensure that these routers are enabled with the following capabilities first:

- LLS (link local signaling)
- OOB (out of band re-synchronization)
- Opaque LSA advertisement
- IETF GR capability

Follow these steps to trigger OSPF Graceful Restart:

To do...	Use the command...	Remarks
Trigger OSPF Graceful Restart	reset ospf [<i>process-id</i>] process graceful-restart	Required Available in user view

Displaying and Maintaining OSPF

To do...	Use the command...	Remarks
Display OSPF brief information	display ospf [<i>process-id</i>] brief	Available in any view
Display OSPF statistics	display ospf [<i>process-id</i>] cumulative	
Display Link State Database information	display ospf [<i>process-id</i>] lsdb [brief [{ ase router network summary asbr nssa opaque-link opaque-area opaque-as } [<i>link-state-id</i>]] [originate-router <i>advertising-router-id</i> self-originate]]	
Display OSPF neighbor information	display ospf [<i>process-id</i>] peer [verbose [<i>interface-type</i> <i>interface-number</i>] [<i>neighbor-id</i>]]	
Display neighbor statistics of OSPF areas	display ospf [<i>process-id</i>] peer statistics	
Display next hop information	display ospf [<i>process-id</i>] nexthop	
Display routing table information	display ospf [<i>process-id</i>] routing [interface <i>interface-type</i> <i>interface-number</i>] [nexthop <i>nexthop-address</i>]	
Display virtual link information	display ospf [<i>process-id</i>] vlink	
Display OSPF request queue information	display ospf [<i>process-id</i>] request-queue [<i>interface-type</i> <i>interface-number</i>] [<i>neighbor-id</i>]	
Display OSPF retransmission queue information	display ospf [<i>process-id</i>] retrans-queue [<i>interface-type</i> <i>interface-number</i>] [<i>neighbor-id</i>]	
Display OSPF ABR and ASBR information	display ospf [<i>process-id</i>] abr-asbr	
Display OSPF interface information	display ospf [<i>process-id</i>] interface [all <i>interface-type</i> <i>interface-number</i>]	
Display OSPF error information	display ospf [<i>process-id</i>] error	
Display OSPF ASBR summarization information	display ospf [<i>process-id</i>] asbr-summary [<i>ip-address</i> { <i>mask</i> <i>mask-length</i> }]	
Reset OSPF counters	reset ospf [<i>process-id</i>] counters [neighbor [<i>interface-type</i> <i>interface-number</i>] [<i>router-id</i>]]	Available in user view
Reset an OSPF process	reset ospf [<i>process-id</i>] process [graceful-restart]	
Remove redistributed routes	reset ospf [<i>process-id</i>] redistribution	

OSPF Configuration Examples



These examples only cover commands for OSPF configuration.

Configuring OSPF Basic Functions

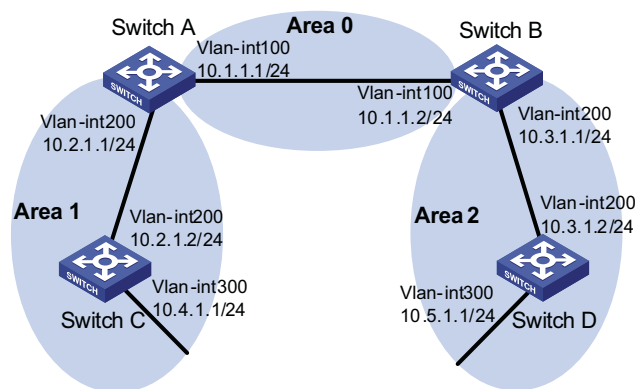
Network requirements

As shown in the following figure, all switches run OSPF. The AS is split into three areas, in which, Switch A and Switch B act as ABRs to forward routing information between areas.

After configuration, all switches can learn routes to every network segment in the AS.

Network diagram

Figure 93 Network diagram for OSPF basic configuration



Configuration procedure

- 1 Configure IP addresses for interfaces (omitted)
- 2 Configure OSPF basic functions

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ospf
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] area 1
[SwitchA-ospf-1-area-0.0.0.1] network 10.2.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.1] quit
[SwitchA-ospf-1] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ospf
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 10.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] area 2
[SwitchB-ospf-1-area-0.0.0.2] network 10.3.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.2] quit
[SwitchB-ospf-1] quit
```

Configure Switch C

```
<SwitchC> system-view
[SwitchC] ospf
[SwitchC-ospf-1] area 1
[SwitchC-ospf-1-area-0.0.0.1] network 10.2.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.1] network 10.4.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.1] quit
[SwitchC-ospf-1] quit
```

Configure Switch D

```
<SwitchD> system-view
[SwitchD] ospf
[SwitchD-ospf-1] area 2
[SwitchD-ospf-1-area-0.0.0.2] network 10.3.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.2] network 10.5.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.2] quit
[SwitchD-ospf-1] quit
```

3 Verify the configuration

Display information about neighbors on Switch A.

```
[SwitchA] display ospf peer verbose

          OSPF Process 1 with Router ID 10.2.1.1
          Neighbors

Area 0.0.0.0 interface 10.1.1.1(Vlan-interface100)'s neighbors
Router ID: 10.3.1.1          Address: 10.1.1.2          GR State: Normal
  State: Full  Mode: Nbr is Master  Priority: 1
  DR: 10.1.1.1  BDR: 10.1.1.2  MTU: 0
  Dead timer due in 37 sec
  Neighbor is up for 06:03:59
  Authentication Sequence: [ 0 ]
  Neighbor state change count: 5

          Neighbors

Area 0.0.0.1 interface 10.2.1.1(Vlan-interface200)'s neighbors
Router ID: 10.4.1.1          Address: 10.2.1.2          GR State: Normal
  State: Full  Mode: Nbr is Master  Priority: 1
  DR: 10.2.1.1  BDR: 10.2.1.2  MTU: 0
  Dead timer due in 32 sec
  Neighbor is up for 06:03:12
  Authentication Sequence: [ 0 ]
  Neighbor state change count: 5
```

Display OSPF routing information on Switch A.

```
[SwitchA] display ospf routing

          OSPF Process 1 with Router ID 10.2.1.1
          Routing Tables

Routing for Network
Destination      Cost      Type      NextHop      AdvRouter      Area
10.2.1.0/24      10        Transit  10.2.1.1      10.2.1.1      0.0.0.1
10.3.1.0/24      4          Inter    10.1.1.2      10.3.1.1      0.0.0.0
10.4.1.0/24      13         Stub    10.2.1.2      10.4.1.1      0.0.0.1
10.5.1.0/24      14         Inter    10.1.1.2      10.3.1.1      0.0.0.0
10.1.1.0/24      2          Transit  10.1.1.1      10.2.1.1      0.0.0.0
```

```
Total Nets: 5
Intra Area: 3 Inter Area: 2 ASE: 0 NSSA: 0
```

Display the Link State Database on Switch A.

```
[SwitchA] display ospf lsdb
```

```
OSPF Process 1 with Router ID 10.2.1.1
Link State Database
```

```
Area: 0.0.0.0
Type      LinkState ID  AdvRouter      Age  Len  Sequence  Metric
Router    10.2.1.1      10.2.1.1       1069 36   80000012   0
Router    10.3.1.1      10.3.1.1       780 36   80000011   0
Network   10.1.1.1      10.2.1.1       1069 32   80000010   0
Sum-Net   10.5.1.0      10.3.1.1       780 28   80000003   12
Sum-Net   10.2.1.0      10.2.1.1       1069 28   8000000F   10
Sum-Net   10.3.1.0      10.3.1.1       780 28   80000014   2
Sum-Net   10.4.1.0      10.2.1.1       769 28   8000000F   13

Area: 0.0.0.1
Type      LinkState ID  AdvRouter      Age  Len  Sequence  Metric
Router    10.2.1.1      10.2.1.1       769 36   80000012   0
Router    10.4.1.1      10.4.1.1      1663 48   80000012   0
Network   10.2.1.1      10.2.1.1       769 32   80000010   0
Sum-Net   10.5.1.0      10.2.1.1       769 28   80000003   14
Sum-Net   10.3.1.0      10.2.1.1       1069 28   8000000F   4
Sum-Net   10.1.1.0      10.2.1.1       1069 28   8000000F   2
Sum-Asbr  10.3.1.1      10.2.1.1       1069 28   8000000F   2
```

Display OSPF routing information on Switch D.

```
[SwitchD] display ospf routing
```

```
OSPF Process 1 with Router ID 10.5.1.1
Routing Tables
```

```
Routing for Network
```

Destination	Cost	Type	NextHop	AdvRouter	Area
10.2.1.0/24	22	Inter	10.3.1.1	10.3.1.1	0.0.0.2
10.3.1.0/24	10	Transit	10.3.1.2	10.3.1.1	0.0.0.2
10.4.1.0/24	25	Inter	10.3.1.1	10.3.1.1	0.0.0.2
10.5.1.0/24	10	Stub	10.5.1.1	10.5.1.1	0.0.0.2
10.1.1.0/24	12	Inter	10.3.1.1	10.3.1.1	0.0.0.2

```
Total Nets: 5
Intra Area: 2 Inter Area: 3 ASE: 0 NSSA: 0
```

On Switch D, ping the IP address 10.4.1.1 to check connectivity.

```
[SwitchD] ping 10.4.1.1
```

```
PING 10.4.1.1: 56 data bytes, press CTRL_C to break
```

```
Request time out
```

```
Reply from 10.4.1.1: bytes=56 Sequence=2 ttl=253 time=15 ms
```

```
Reply from 10.4.1.1: bytes=56 Sequence=3 ttl=253 time=1 ms
```

```
Reply from 10.4.1.1: bytes=56 Sequence=4 ttl=253 time=16 ms
```

```
Reply from 10.4.1.1: bytes=56 Sequence=5 ttl=253 time=1 ms
```

```
--- 10.4.1.1 ping statistics ---
```

```
5 packet(s) transmitted
```

```
4 packet(s) received
```

```
20.00% packet loss
```

```
round-trip min/avg/max = 1/8/16 ms
```

Configuring an OSPF Stub Area

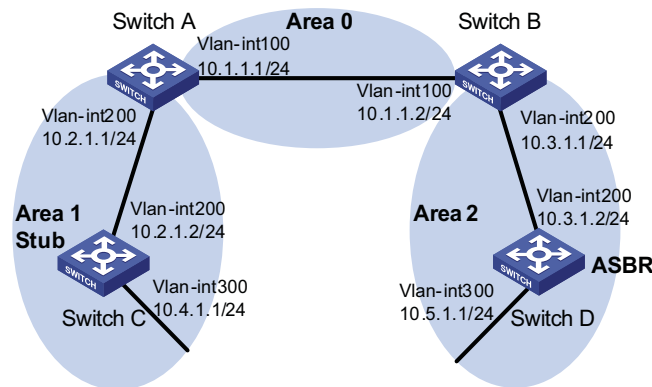
Network requirements

The following figure shows an AS is split into three areas, where all switches run OSPF. Switch A and Switch B act as ABRs to forward routing information between areas. Switch D acts as the ASBR to redistribute routes (static routes).

It is required to configure Area 1 as a Stub area, reducing LSAs to this area without affecting route reachability.

Network diagram

Figure 94 Network diagram for OSPF Stub area configuration



- 1 Configure IP addresses for interfaces (omitted).
- 2 Configure OSPF basic functions (refer to “Configuring OSPF Basic Functions” on page 310).
- 3 Configure Switch D to redistribute static routes.

```
[SwitchD] ip route-static 3.1.2.1 24 10.5.1.2
[SwitchD] ospf
[SwitchD-ospf-1] import-route static
[SwitchD-ospf-1] quit
```

Display ABR/ASBR information on Switch C.

```
[SwitchC] display ospf abr-asbr
```

```
OSPF Process 1 with Router ID 10.4.1.1
Routing Table to ABR and ASBR
```

Type	Destination	Area	Cost	NextHop	RtType
Intra	10.2.1.1	0.0.0.1	3	10.2.1.1	ABR
Inter	10.3.1.1	0.0.0.1	5	10.2.1.1	ABR
Inter	10.5.1.1	0.0.0.1	7	10.2.1.1	ASBR

Display OSPF routing table information on Switch C.

```
[SwitchC] display ospf routing
```

```
OSPF Process 1 with Router ID 10.4.1.1
Routing Tables
```

Routing for Network					
Destination	Cost	Type	NextHop	AdvRouter	Area
10.2.1.0/24	3	Transit	10.2.1.2	10.2.1.1	0.0.0.1
10.3.1.0/24	7	Inter	10.2.1.1	10.2.1.1	0.0.0.1
10.4.1.0/24	3	Stub	10.4.1.1	10.4.1.1	0.0.0.1

```

10.5.1.0/24      17      Inter  10.2.1.1      10.2.1.1      0.0.0.1
10.1.1.0/24     5       Inter  10.2.1.1      10.2.1.1      0.0.0.1

```

```

Routing for ASEs
Destination      Cost      Type      Tag      NextHop      AdvRouter
3.1.2.0/24      1         Type2     1         10.2.1.1     10.5.1.1

```

```

Total Nets: 6
Intra Area: 2  Inter Area: 3  ASE: 1  NSSA: 0

```



In the above output, since Switch C resides in a normal OSPF area, its routing table contains an external route.

4 Configure Area 1 as a Stub area.

Configure Switch A.

```

[SwitchA] ospf
[SwitchA-ospf-1] area 1
[SwitchA-ospf-1-area-0.0.0.1] stub
[SwitchA-ospf-1-area-0.0.0.1] quit
[SwitchA-ospf-1] quit

```

Configure Switch C.

```

[SwitchC] ospf
[SwitchC-ospf-1] area 1
[SwitchC-ospf-1-area-0.0.0.1] stub
[SwitchC-ospf-1-area-0.0.0.1] quit
[SwitchC-ospf-1] quit

```

Display OSPF routing information on Switch C

```

[SwitchC] display ospf routing

          OSPF Process 1 with Router ID 10.4.1.1
          Routing Tables

Routing for Network
Destination      Cost      Type      NextHop      AdvRouter      Area
0.0.0.0/0       4         Inter     10.2.1.1     10.2.1.1       0.0.0.1
10.2.1.0/24     3         Transit  10.2.1.2     10.2.1.1       0.0.0.1
10.3.1.0/24     7         Inter     10.2.1.1     10.2.1.1       0.0.0.1
10.4.1.0/24     3         Stub     10.4.1.1     10.4.1.1       0.0.0.1
10.5.1.0/24     17        Inter     10.2.1.1     10.2.1.1       0.0.0.1
10.1.1.0/24     5         Inter     10.2.1.1     10.2.1.1       0.0.0.1

Total Nets: 6
Intra Area: 2  Inter Area: 4  ASE: 0  NSSA: 0

```



When Switch C resides in the Stub area, a default route takes the place of the external route.

Filter Type-3 LSAs out the stub area

```

[SwitchA] ospf
[SwitchA-ospf-1] area 1
[SwitchA-ospf-1-area-0.0.0.1] stub no-summary
[SwitchA-ospf-1-area-0.0.0.1] quit

```

Display OSPF routing information on Switch C.

```
[SwitchC] display ospf routing
```

```
OSPF Process 1 with Router ID 10.4.1.1
Routing Tables
```

```
Routing for Network
```

Destination	Cost	Type	NextHop	AdvRouter	Area
0.0.0.0/0	4	Inter	10.2.1.1	10.2.1.1	0.0.0.1
10.2.1.0/24	3	Transit	10.2.1.2	10.4.1.1	0.0.0.1
10.4.1.0/24	3	Stub	10.4.1.1	10.4.1.1	0.0.0.1

```
Total Nets: 3
```

```
Intra Area: 2 Inter Area: 1 ASE: 0 NSSA: 0
```



After this configuration, routing entries on the stub router are further reduced, containing only one default external route.

Configuring an OSPF NSSA Area

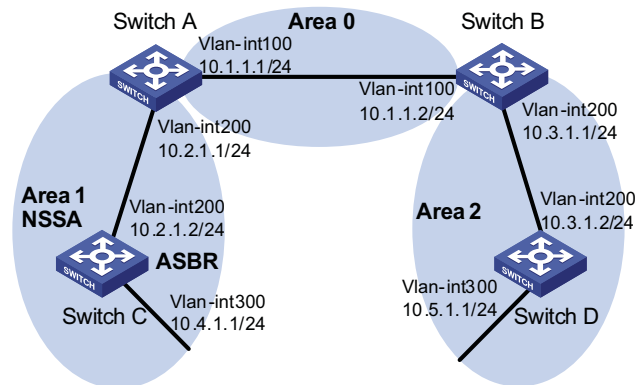
Network requirements

The following figure shows an AS is split into three areas, where all switches run OSPF. Switch A and Switch B act as ABRs to forward routing information between areas.

It is required to configure Area 1 as an NSSA area, and configure Router C as the ASBR to redistribute static routes into the AS.

Network diagram

Figure 95 Network diagram for OSPF NSSA area configuration



Configuration procedure

- 1 Configure IP addresses for interfaces.
- 2 Configure OSPF basic functions (refer to “Configuring OSPF Basic Functions” on page 310).
- 3 Configure Area 1 as an NSSA area.

Configure Switch A.

```
[SwitchA] ospf
[SwitchA-ospf-1] area 1
[SwitchA-ospf-1-area-0.0.0.1] nssa default-route-advertise no-summary
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
```

Configure Switch C.

```
[SwitchC] ospf
[SwitchC-ospf-1] area 1
[SwitchC-ospf-1-area-0.0.0.1] nssa
[SwitchC-ospf-1-area-0.0.0.1] quit
[SwitchC-ospf-1] quit
```



*It is recommended to configure the **nssa** command with the keyword **default-route-advertise no-summary** on Switch A (an ABR) to reduce the routing table size on NSSA routers. On other NSSA routers, using the **nssa** command is ok.*

Display OSPF routing information on Switch C.

```
[SwitchC] display ospf routing

          OSPF Process 1 with Router ID 10.4.1.1
          Routing Tables

Routing for Network
Destination      Cost      Type      NextHop      AdvRouter      Area
0.0.0.0/0        65536     Inter     10.2.1.1     10.2.1.1       0.0.0.1
10.2.1.0/24      65535     Transit   10.2.1.2     10.4.1.1       0.0.0.1
10.4.1.0/24      3         Stub      10.4.1.1     10.4.1.1       0.0.0.1

Total Nets: 3
Intra Area: 2  Inter Area: 1  ASE: 0  NSSA: 0
```

4 Configure Switch C to redistribute static routes.

```
[SwitchC] ip route-static 3.1.3.1 24 11.1.1.1
[SwitchC] ospf
[SwitchC-ospf-1] import-route static
[SwitchC-ospf-1] quit
```

Display OSPF routing information on Switch D.

```
[SwitchD-ospf-1] display ospf routing

          OSPF Process 1 with Router ID 10.5.1.1
          Routing Tables

Routing for Network
Destination      Cost      Type      NextHop      AdvRouter      Area
10.2.1.0/24      22        Inter     10.3.1.1     10.3.1.1       0.0.0.2
10.3.1.0/24      10        Transit   10.3.1.2     10.3.1.1       0.0.0.2
10.4.1.0/24      25        Inter     10.3.1.1     10.3.1.1       0.0.0.2
10.5.1.0/24      10        Stub      10.5.1.1     10.5.1.1       0.0.0.2
10.1.1.0/24      12        Inter     10.3.1.1     10.3.1.1       0.0.0.2

Routing for ASEs
Destination      Cost      Type      Tag      NextHop      AdvRouter
3.1.3.0/24       1         Type2     1         10.3.1.1     10.2.1.1

Total Nets: 6
Intra Area: 2  Inter Area: 3  ASE: 1  NSSA: 0
```



You can see on Switch D an external route imported from the NSSA area.

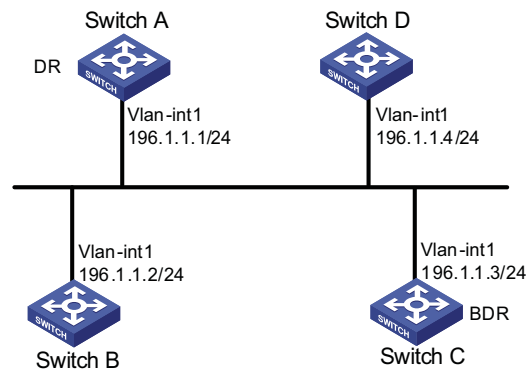
Configuring OSPF DR Election

Network requirements

- In the following figure, OSPF Switches A, B, C and D reside on the same network segment.
- It is required to configure Switch A as the DR, and configure Switch C as the BDR.

Network diagram

Figure 96 Network diagram for OSPF DR election configuration



Configuration procedure

- 1 Configure IP addresses for interfaces (omitted)
- 2 Configure OSPF basic functions

Configure Switch A.

```
<SwitchA> system-view
[Switch A] router id 1.1.1.1
[Switch A] ospf
[Switch A-ospf-1] area 0
[Switch A-ospf-1-area-0.0.0.0] network 196.1.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] router id 2.2.2.2
[SwitchB] ospf
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 196.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] router id 3.3.3.3
[SwitchC] ospf
[SwitchC-ospf-1] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 196.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit
```

Configure Switch D.

```

<SwitchD> system-view
[SwitchD] router id 4.4.4.4
[SwitchD] ospf
[SwitchD-ospf-1] area 0
[SwitchD-ospf-1-area-0.0.0.0] network 196.1.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] quit
[SwitchD-ospf-1] quit

```

Display OSPF neighbor information on Switch A.

```

[SwitchA] display ospf peer verbose
      OSPF Process 1 with Router ID 1.1.1.1
      Neighbors
Area 0.0.0.0 interface 192.168.1.1(Vlan-interface1)'s neighbors
Router ID: 2.2.2.2      Address: 192.168.1.2      GR State: Normal
  State: 2-Way  Mode: None  Priority: 1
  DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
  Dead timer due in 38 sec
  Neighbor is up for 00:01:31
  Authentication Sequence: [ 0 ]
  Neighbor state change count: 2

Router ID: 3.3.3.3      Address: 192.168.1.3      GR State: Normal
  State: Full  Mode: Nbr is Master  Priority: 1
  DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
  Dead timer due in 31 sec
  Neighbor is up for 00:01:28
  Authentication Sequence: [ 0 ]
  Neighbor state change count: 2

Router ID: 4.4.4.4      Address: 192.168.1.4      GR State: Normal
  State: Full  Mode: Nbr is Master  Priority: 1
  DR: 192.168.1.4  BDR: 192.168.1.3  MTU: 0
  Dead timer due in 31 sec
  Neighbor is up for 00:01:28
  Authentication Sequence: [ 0 ]
  Neighbor state change count: 2

```

Switch D becomes the DR, and Switch C is the BDR.

3 Configure router priorities on interfaces**# Configure Switch A.**

```

[SwitchA] interface vlan-interface 1
[RouterA-Vlan-interface1] ospf dr-priority 100
[RouterA-Vlan-interface1] quit

```

Configure Switch B.

```

[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ospf dr-priority 0
[SwitchB-Vlan-interface1] quit

```

Configure Switch C.

```

[SwitchC] interface vlan-interface 1
[SwitchC-Vlan-interface1] ospf dr-priority 2
[SwitchC-Vlan-interface] quit

```

Display neighbor information on Switch D.

```
[SwitchD] display ospf peer verbose
      OSPF Process 1 with Router ID 4.4.4.4
      Neighbors
Area 0.0.0.0 interface 192.168.1.4(Vlan-interface1)'s neighbors
Router ID: 1.1.1.1      Address: 192.168.1.1      GR State: Normal
  State: Full Mode:Nbr is Slave Priority: 100
  DR: 192.168.1.4 BDR: 192.168.1.3 MTU: 0
  Dead timer due in 31 sec
  Neighbor is up for 00:11:17
  Authentication Sequence: [ 0 ]
  Neighbor state change count: 5

Router ID: 2.2.2.2      Address: 192.168.1.2      GR State: Normal
  State: Full Mode:Nbr is Slave Priority: 0
  DR: 192.168.1.4 BDR: 192.168.1.3 MTU: 0
  Dead timer due in 35 sec
  Neighbor is up for 00:11:19
  Authentication Sequence: [ 0 ]
  Neighbor state change count: 5

Router ID: 3.3.3.3      Address: 192.168.1.3      GR State: Normal
  State: Full Mode:Nbr is Slave Priority: 2
  DR: 192.168.1.4 BDR: 192.168.1.3 MTU: 0
  Dead timer due in 33 sec
  Neighbor is up for 00:11:15
  Authentication Sequence: [ 0 ]
  Neighbor state change count: 5
```

The DR and BDR have no change.



In the above output, you can find the priority configuration does not take effect immediately.

4 Restart OSPF process (omitted)

Display neighbor information on Switch D.

```
[SwitchD] display ospf peer verbose
      OSPF Process 1 with Router ID 4.4.4.4
      Neighbors
Area 0.0.0.0 interface 192.168.1.4(Vlan-interface1)'s neighbors
Router ID: 1.1.1.1      Address: 192.168.1.1      GR State: Normal
  State: Full Mode: Nbr is Slave Priority: 100
  DR: 192.168.1.1 BDR: 192.168.1.3 MTU: 0
  Dead timer due in 39 sec
  Neighbor is up for 00:01:40
  Authentication Sequence: [ 0 ]
  Neighbor state change count: 2

Router ID: 2.2.2.2      Address: 192.168.1.2      GR State: Normal
  State: 2-Way Mode: None Priority: 0
  DR: 192.168.1.1 BDR: 192.168.1.3 MTU: 0
  Dead timer due in 35 sec
  Neighbor is up for 00:01:44
  Authentication Sequence: [ 0 ]
  Neighbor state change count: 2

Router ID: 3.3.3.3      Address: 192.168.1.3      GR State: Normal
  State: Full Mode: Nbr is Slave Priority: 2
```

```
DR: 192.168.1.1 BDR: 192.168.1.3 MTU: 0
Dead timer due in 39 sec
Neighbor is up for 00:01:41
Authentication Sequence: [ 0 ]
Neighbor state change count: 2
```

Switch A becomes the DR, and Switch C is the BDR.



If the neighbor state is full, it means Switch D has established the adjacency with the neighbor. If the neighbor state is 2-way, it means the two switches are neither the DR nor the BDR, and they do not exchange LSAs.

Display OSPF interface information.

```
[SwitchA] display ospf interface
```

```
OSPF Process 1 with Router ID 1.1.1.1
Interfaces
```

```
Area: 0.0.0.0
IP Address      Type      State    Cost  Pri  DR          BDR
192.168.1.1     Broadcast DR       1      100  192.168.1.1 192.168.1.3
```

```
[SwitchB] display ospf interface
```

```
OSPF Process 1 with Router ID 2.2.2.2
Interfaces
```

```
Area: 0.0.0.0
IP Address      Type      State    Cost  Pri  DR          BDR
192.168.1.2     Broadcast DROther 1      0     192.168.1.1 192.168.1.3
```



The interface state DROther means the interface is not the DR/BDR.

Configuring OSPF Virtual Links

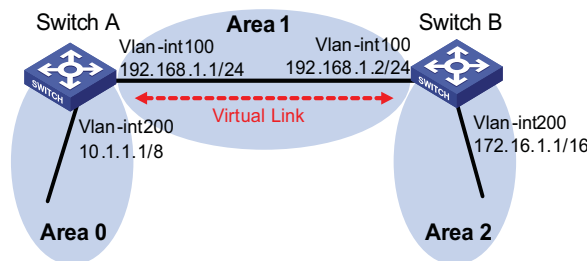
Network requirements

In Figure 97, Area 2 has no direct connection to Area 0, and Area 1 acts as the Transit Area to connect Area 2 to Area 0 via a configured virtual link between Switch B and Switch C.

After configuration, Switch A can learn routes to Area 2.

Network diagram

Figure 97 Network diagram for OSPF virtual link configuration



Configuration procedure

- 1 Configure IP addresses for interfaces (omitted)
- 2 Configure OSPF basic functions

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ospf 1 router-id 1.1.1.1
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 10.0.0.0 0.255.255.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] area 1
[SwitchA-ospf-1-area-0.0.0.1] network 192.168.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.1] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ospf 1 router-id 2.2.2.2
[SwitchB-ospf-1] area 1
[SwitchB-ospf-1-area-0.0.0.1] network 192.168.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.1] quit
[SwitchB-ospf-1] area 2
[SwitchB-ospf-1-area-0.0.0.2] network 172.16.0.0 0.0.255.255
[SwitchB-ospf-1-area-0.0.0.2] quit
```

Display OSPF routing information on Switch A.

```
[SwitchA] display ospf routing
      OSPF Process 1 with Router ID 1.1.1.1
      Routing Tables

Routing for Network
Destination      Cost      Type      NextHop      AdvRouter      Area
10.0.0.0/8       1         Stub     10.1.1.1     1.1.1.1        0.0.0.0
192.168.1.0/24  1562     Stub     192.168.1.1 1.1.1.1        0.0.0.1

Total Nets: 2
Intra Area: 2  Inter Area: 0  ASE: 0  NSSA: 0
```



Since Area 2 has no direct connection to Area 0, the OSPF routing table of Router A has no route to Area 2.

3 Configure a virtual link

Configure Switch A.

```
[SwitchA] ospf
[SwitchA-ospf-1] area 1
[SwitchA-ospf-1-area-0.0.0.1] vlink-peer 2.2.2.2
[SwitchA-ospf-1-area-0.0.0.1] quit
[SwitchA-ospf-1] quit
```

Configure Switch B.

```
[SwitchB] ospf 1
[SwitchB-ospf-1] area 1
[SwitchB-ospf-1-area-0.0.0.1] vlink-peer 1.1.1.1
[SwitchB-ospf-1-area-0.0.0.1] quit
```

Display OSPF routing information on Switch A.

```
[SwitchA] display ospf routing

                OSPF Process 1 with Router ID 1.1.1.1
                Routing Tables

Routing for Network
Destination      Cost  Type      NextHop      AdvRouter     Area
172.16.1.1/16   1563  Inter     192.168.1.2  2.2.2.2       0.0.0.0
10.0.0.0/8      1     Stub     10.1.1.1     1.1.1.1       0.0.0.0
192.168.1.0/24  1562  Stub     192.168.1.1  1.1.1.1       0.0.0.1

Total Nets: 3
Intra Area: 2  Inter Area: 1  ASE: 0  NSSA: 0
```

Switch A has learned the route 172.16.1.1/16 to Area 2.

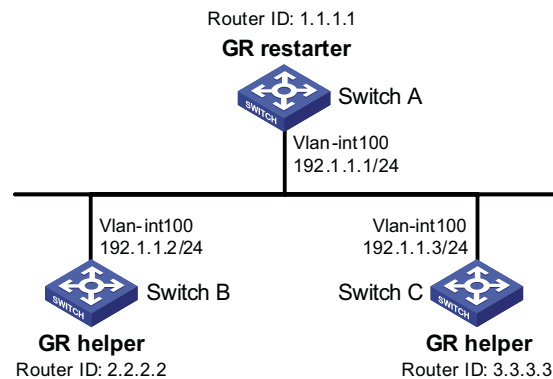
OSPF Graceful Restart Configuration Example

Network requirements

- Switch A, Switch B and Switch C that belong to the same autonomous system and the same OSPF routing domain are GR capable.
- Switch A acts as the non IETF standard GR Restarter whereas Switch B and Switch C are the GR Helpers and remain OOB synchronized with Switch A through the GR mechanism.

Network diagram

Figure 98 Network diagram for OSPF-based GR configuration



Configuration procedure

1 Configure Switch A

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ip address 192.1.1.1 255.255.255.0
[SwitchA-Vlan-interface100] quit
[SwitchA] router id 1.1.1.1
[SwitchA] ospf 100
[SwitchA-ospf-100] enable link-local-signaling
[SwitchA-ospf-100] enable out-of-band-resynchronization
[RouterA-ospf-100] graceful-restart
[SwitchA-ospf-100] area 0
[SwitchA-ospf-100-area-0.0.0.0] network 192.1.1.0 0.0.0.255
[SwitchA-ospf-100-area-0.0.0.0] return
```

2 Configure Switch B

```

<SwitchB> system-view
[SwitchB] acl number 2000
[SwitchB-acl-basic-2000] rule 10 permit source 192.1.1.1 0.0.0.0
[SwitchB-acl-basic-2000] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ip address 192.1.1.2 255.255.255.0
[SwitchB-Vlan-interface100] ospf dr-priority 0
[SwitchB-Vlan-interface100] quit
[SwitchB] router id 2.2.2.2
[SwitchB] ospf 100
[SwitchB-ospf-100] enable link-local-signaling
[SwitchB-ospf-100] enable out-of-band-resynchronization
[SwitchB-ospf-100] graceful-restart help 2000
[SwitchB-ospf-100] area 0
[SwitchB-ospf-100-area-0.0.0.0] network 192.1.1.0 0.0.0.255
[SwitchB-ospf-100-area-0.0.0.0] quit

```

3 Configure Switch C

```

<SwitchC> system-view
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] ip address 192.1.1.3 255.255.255.0
[SwitchC-Vlan-interface100] ospf dr-priority 2
[SwitchC-Vlan-interface100] quit
[SwitchC] router id 3.3.3.3
[SwitchC] ospf 100
[SwitchC-ospf-100] enable link-local-signaling
[SwitchC-ospf-100] enable out-of-band-resynchronization
[SwitchC-ospf-100] area 0
[SwitchC-ospf-100-area-0.0.0.0] network 192.1.1.0 0.0.0.255
[SwitchC-ospf-100-area-0.0.0.0] quit

```

4 Verify the configuration

After the configurations on Switch A, Switch B and Switch C are completed and the switches are running steadily, perform OSPF GR on Switch A.

```

<SwitchA> reset ospf 100 process graceful-restart

```

Troubleshooting OSPF Configuration

No OSPF Neighbor Relationship Established

Symptom

No OSPF neighbor relationship can be established.

Analysis

If the physical link and lower layer protocols work well, check OSPF parameters configured on interfaces. Two neighbors must have the same parameters, such as the area ID, network segment and mask (a P2P or virtual link may have different network segments and masks), network type. If the network type is broadcast or NBMA, at least one interface must have a router priority higher than 0.

Processing steps

- 1 Display OSPF neighbor information using the **display ospf peer** command.
- 2 Display OSPF interface information using the **display ospf interface** command.

- 3 Ping the neighbor router's IP address to check connectivity.
- 4 Check OSPF timers. The neighbor dead interval on an interface must be at least four times the hello interval.
- 5 On an NBMA network, using the **peer ip-address** command to specify the neighbor manually is required.
- 6 On an NBMA or a broadcast network, at least one connected interface must have a router priority higher than 0.

Incorrect Routing Information

Symptom

OSPF cannot find routes to other areas.

Analysis

The backbone area must maintain connectivity to all other areas. If a router connects to more than one area, at least one area must be connected to the backbone. The backbone cannot be configured as a Stub area.

In a Stub area, all routers cannot receive external routes, and all interfaces connected to the Stub area must belong to the Stub area.

Solution

- 1 Use the **display ospf peer** command to display neighbors.
- 2 Use the **display ospf interface** command to display OSPF interface information.
- 3 Use the **display ospf lsdb** command to display the Link State Database to check its integrity.
- 4 Display information about area configuration using the **display current-configuration configuration ospf** command. If more than two areas are configured, at least one area is connected to the backbone.
- 5 In a Stub area, all routers attached are configured with the **stub** command. In an NSSA area, all interface connected to which are configured with the **nssa** command.
- 6 If a virtual link is configured, use the **display ospf vlink** command to check the state of the virtual link.

29

IS-IS CONFIGURATION

When configuring IS-IS, go to these sections for information you are interested in:

- "IS-IS Overview" on page 325
- "IS-IS Configuration Task List" on page 340
- "Configuring IS-IS Basic Functions" on page 341
- "Configuring IS-IS Routing Information Control" on page 342
- "Tuning and Optimizing IS-IS Network" on page 346
- "Configuring IS-IS GR" on page 352
- "Displaying and Maintaining IS-IS" on page 353
- "IS-IS Configuration Example" on page 354



The term "router" in this document refers to a router in a generic sense or an Ethernet switch running routing protocols.

IS-IS Overview

Intermediate System-to-Intermediate System (IS-IS) is a dynamic routing protocol designed by the International Organization for Standardization (ISO) to operate on the connectionless network protocol (CLNP).

The IS-IS routing protocol has been modified and extended in RFC 1195 by the International Engineer Task Force (IETF) for application in both TCP/IP and OSI reference models, and the new one is called Integrated IS-IS or Dual IS-IS.

IS-IS is an interior gateway protocol (IGP) used within an Autonomous System. It adopts the Shortest Path First (SPF) algorithm for route calculation.

Basic Concepts **IS-IS terminology**

- Intermediate system (IS). An IS, similar to a router in TCP/IP, is the basic unit in IS-IS protocol to generate and propagate routing information. In the following text, an IS is a router.
- End system (ES). An ES refers to a host system in TCP/IP. ISO defines the ES-IS protocol for communication between an ES and an IS, therefore an ES does not participate in the IS-IS process.
- Routing domain (RD). A group of ISs exchange routing information with the same routing protocol in a routing domain.
- Area. An area is a division unit in a routing domain. The IS-IS protocol allows a routing domain to be divided into multiple areas.

- Link State Database (LSDB). All link states in the network forms the LSDB. There is at least one LSDB in each IS. The IS uses SPF algorithm and LSDB to generate its own routes.
- Link State Protocol Data Unit (LSPDU) or Link State Packet (LSP). Each IS can generate a LSP which contains all the link state information of the IS. Each IS collects all the LSPs in the local area to generate its own LSDB.
- Network Protocol Data Unit (NPDU). An NPDU is a network layer protocol packet in ISO, which is equivalent to an IP packet in TCP/IP.
- Designated IS. On a broadcast network, the designated router is also known as the designated IS or a pseudonode.
- Network service access point (NSAP). The NSAP is the ISO network layer address. It identifies an abstract network service access point and describes the network address in the ISO reference model.

IS-IS address structure

1 NSAP

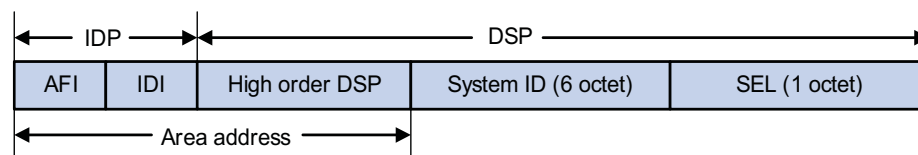
As shown in Figure 99, the NSAP address consists of the Initial Domain Part (IDP) and the Domain Specific Part (DSP). The IDP is equal to the network ID of the IP address, and the DSP is equal to the subnet and host IDs.

The IDP, defined by ISO, includes the Authority and Format Identifier (AFI) and the Initial Domain Identifier (IDI).

The DSP includes the High Order DSP (HODSP), the System ID and SEL, where the HODSP identifies the area, the System ID identifies the host, and the SEL indicates the type of service.

The length of IDP and DSP is variable. The length of the NSAP address varies from 8 bytes to 20 bytes.

Figure 99 NSAP address structure



2 Area address

The area address is composed of the IDP and the HODSP of the DSP, which identify the area and the routing domain. Different routing domains cannot have the same area address.

Generally, a router only needs one area address, and all nodes in the same routing domain must share the same area address. However, a router can have three area addresses at most to support smooth area merging, partitioning and switching.

3 System ID

The system ID identifies the host or router uniquely. It has a fixed length of 48 bits (6 bytes).

The system ID is used in cooperation with the Router ID in practical. For example, a router uses the IP address 168.10.1.1 of the Loopback 0 as the Router ID, the system ID in IS-IS can be obtained in the following way:

- Extend each decimal number of the IP address to 3 digits by adding 0s from the left, like 168.010.001.001;
- Divide the extended IP address into 3 sections with 4 digits in each section to get the System ID 1680.1000.1001.

There are other methods to define a system ID. Just make sure it can uniquely identify a host or router.

4 SEL

The NSAP Selector (SEL), sometimes present in N-SEL, is similar with the protocol identifier in IP. Different transport layer protocols use different SELs. All SELs in IP are 00.

5 Routing method

Since the area is explicitly defined in the address structure, the Level-1 router can easily recognize the packets sent out of the area. These packets are forwarded to the Level-2 router.

The Level-1 router makes routing decisions based on the system ID. If the destination is not in the area, the packet is forwarded to the nearest Level-1-2 router.

The Level-2 router routes packets across areas according to the area address.

NET

The Network Entity Title (NET) is an NSAP with SEL of 0. It indicates the network layer information of the IS itself, where SEL=0 means no transport layer information. Therefore, the length of NET is equal to NSAP, in the range 8 bytes to 20 bytes.

Generally, a router only needs one NET, but it can have three NETs at most for smooth area merging and partitioning. When you configure multiple NETs, make sure their system IDs are the same.

For example, a NET is ab.cdef.1234.5678.9abc.00, where,

Area = ab.cdef, System ID = 1234.5678.9abc, and SEL = 00.

IS-IS Area Two-level hierarchy

IS-IS uses two-level hierarchy in the routing domain to support large scale routing networks. A large routing domain is divided into multiple Areas. The Level-1 router is in charge of forwarding routes within an area, and the Level-2 router is in charge of forwarding routes between areas.

Level-1 and Level-2

1 Level-1 router

The Level-1 router only establishes the neighbor relationship with Level-1 and Level-1-2 routers in the same area. The LSDB maintained by the Level-1 router contains the local area routing information. It directs the packets out of the area to the nearest Level-1-2 router.

2 Level-2 router

The Level-2 router establishes the neighbor relationships with the Level-2 and Level-1-2 routers in the same or in different areas. It maintains a Level-2 LSDB which contains inter area routing information. All the Level-2 and Level-1-2 routers must be contiguous to form the backbone in a routing domain. Only Level-2 routers can directly communicate with routers outside the routing domain.

3 Level-1-2 router

A router with both Level-1 and Level-2 router functions is called a Level-1-2 router. It can establish the Level-1 neighbor relationship with the Level-1 and Level-1-2 routers in the same area, or establish Level-2 neighbor relationship with the Level-2 and Level-1-2 routers in different areas. A Level-1 router must be connected to other areas via a Level-1-2 router. The Level-1-2 router maintains two LSDBs, where the Level-1 LSDB is for routing within the area, and the Level-2 LSDB is for routing between areas.



- *The Level-1 routers in different areas can not establish the neighbor relationship.*
- *The neighbor relationship establishment of Level-2 routers has nothing to do with area.*

Figure 100 shows a network topology running the IS-IS protocol. Area 1 is a set of Level-2 routers, called backbone network. The other four areas are non-backbone networks connected to the backbone through Level-1-2 routers.

Figure 100 IS-IS topology

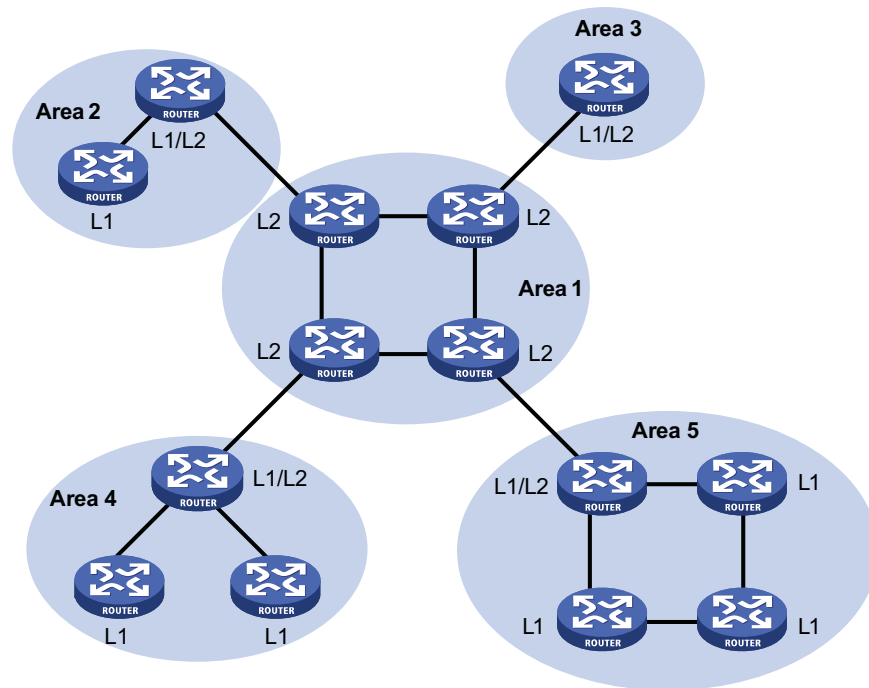
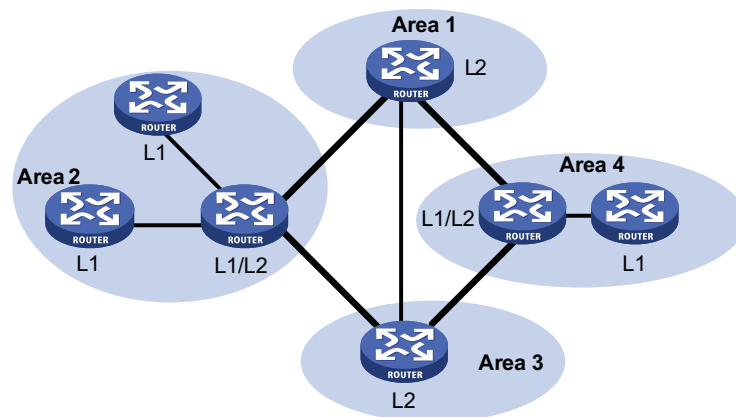


Figure 101 shows another network topology running the IS-IS protocol. The Level-1-2 routers connect the Level-1 and Level-2 routers, and also form the IS-IS backbone together with the Level-2 routers. There is no area defined as the backbone in this topology. The backbone is composed of all contiguous Level-2 and Level-1-2 routers which can reside in different areas.

Figure 101 IS-IS topology



The IS-IS backbone does not need to be a specific Area.

Both the IS-IS Level-1 and Level-2 routers use the SPF algorithm to generate the Shortest Path Tree (SPT).

Interface routing hierarchy type

You can configure the routing type for each interface. For a Level-1-2 router, one interface may establish Level-1 adjacency with a router, and another one may establish Level-2 adjacency with another router. You can limit the adjacency type

by configuring the routing hierarchy on the interface. For example, the level-1 interface can only establish Level-1 adjacency, while the level-2 interface can only establish Level-2 adjacency.

By having this function, you can prevent the Level-1 hello packets from propagating to the Level-2 backbone through the Level-1-2 router. This can result in bandwidth saving.

Route leaking

An IS-IS routing domain is comprised of only one Level-2 area and multiple Level-1 areas. A Level-1 area is connected with the Level-2 area rather than other Level-1 areas.

The routing information of the Level-1 area is sent to the Level-2 area through the Level-1-2 router. Therefore, the Level-2 router knows the routing information of the entire IS-IS routing domain but does not share the information with the Level-1 area by default.

Since the Level-1 router simply sends the routing information for destinations outside the area to the nearest Level-1-2 router, this may cause a problem that the best path cannot be selected.

To solve this problem, route leaking was introduced. The Level-2 router can advertise the Level-2 routing information to a specified Level-1 area. By having the routing information of other areas, the Level-1 router can make a better routing choice for the packets destined outside the area.

IS-IS Network Type **Network type**

IS-IS supports two network types:

- Broadcast network, such as Ethernet, Token-Ring.
- Point-to-point network, such as PPP, HDLC.



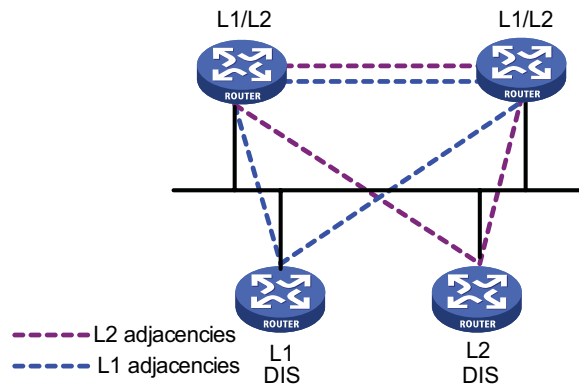
For the Non-Broadcast Multi-Access (NBMA) network, such as ATM, you need to configure point-to-point or broadcast network on its configured subinterfaces. IS-IS does not run on Point to Multipoint (P2MP) links.

DIS and pseudo nodes

On an IS-IS broadcast network, a router has to be selected as the Designated Intermediate System (DIS).

The Level-1 and Level-2 DISs are selected respectively. You can assign different priorities for different level DIS selections. The higher a router's priority is, the more likelihood the router becomes the DIS. If there are multiple routers with the same highest DIS priority, the one with the highest SNPA (Subnetwork Point of Attachment) address (which is the MAC address on a broadcast network) will be selected. A router can be the DIS for different levels.

As shown in Figure 102, the same level routers on the same network segment can establish adjacencies. This is different from OSPF.

Figure 102 DIS in the IS-IS broadcast network

The DIS creates and updates pseudo nodes as well as their LSP to describe all routers on the network.

The pseudonode emulates a virtual node on the broadcast network. It is not a real router. In IS-IS, it is identified by the system ID and one byte Circuit ID (a non zero value) of the DIS.

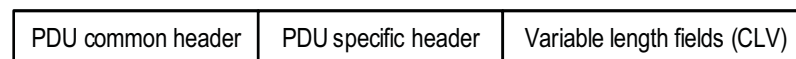
Using pseudo nodes can reduce LSPs, the resources used by SPF and simplify the network topology.



On IS-IS broadcast networks, all routers are adjacent with each other. The DIS is responsible for the synchronization of their LSDBs.

IS-IS PDU Format PDU header format

The IS-IS packets are encapsulated into link layer frames. The Protocol Data Unit (PDU) consists of two parts, the headers and the variable length field, where the headers can be further divided into the common header and the specific header. The common headers are the same for all PDUs, while the specific headers vary by PDU type. The following figure shows the PDU format.

Figure 103 PDU format

Common header format

Figure 104 shows the common header format.

Figure 104 PDU common header format

				No. of Octets
Intradomain routing protocol discriminator				1
Length indicator				1
Version/Protocol ID extension				1
ID length				1
R	R	R	PDU type	1
Version				1
Reserved				1
Maximum area address				1

- Intra-domain Routing Protocol Discriminator: Set to 0x83.
- Length Indicator: The length of the PDU header, including both common and specific headers, present in bytes.
- Version/Protocol ID Extension: Set to 1(0x01).
- ID Length: The length of the NSAP address and NET ID.
- R (Reserved): Set to 0.
- PDU Type: For detail information, refer to Table 44.
- Version: Set to 1(0x01).
- Maximum Area Address: Maximum number of area addresses supported.

Table 44 PDU type

Type	PDU Type	Acronym
15	Level-1 LAN IS-IS hello PDU	L1 LAN IIH
16	Level-2 LAN IS-IS hello PDU	L2 LAN IIH
17	Point-to-Point IS-IS hello PDU	P2P IIH
18	Level-1 Link State PDU	L1 LSP
20	Level-2 Link State PDU	L2 LSP
24	Level-1 Complete Sequence Numbers PDU	L1 CSNP
25	Level-2 Complete Sequence Numbers PDU	L2 CSNP
26	Level-1 Partial Sequence Numbers PDU	L1 PSNP
27	Level-2 Partial Sequence Numbers PDU	L2 PSNP

Hello

The hello packet is used by routers to establish and maintain the neighbor relationship. It is also called IS-to-IS hello PDU (IIH). For broadcast network, the Level-1 router uses the Level-1 LAN IIH; and the Level-2 router uses the Level-2 LAN IIH. The P2P IIH is used on point-to-point network.

Figure 105 illustrates the hello packet format in broadcast networks, where the blue fields are the common header.

Figure 105 L1/L2 LAN IIH format

				No. of Octets
Intradomain routing protocol discriminator				1
Length indicator				1
Version/Protocol ID extension				1
ID length				1
R	R	R	PDU type	1
Version				1
Reserved				1
Maximum area address				1
Reserved/Circuit type				1
Source ID				ID length
Holding time				2
PDU length				2
R	Priority			1
LAN ID				ID length+ 1
Variable length fields				

- Reserved/Circuit Type: The first 6 bits are reserved with value 0. The last 2 bits indicates router types: 00 means reserved, 01 indicates L1, 10 indicates L2, and 11 indicates L1/2.
- Source ID: The system ID of the router advertising the hello packet.
- Holding Time: If no hello packets are received from a neighbor within the holding time, the neighbor is considered dead.
- PDU Length: The total length of the PDU in bytes.
- Priority: DIS priority.
- LAN ID: Includes the system ID and one byte pseudonode ID.

Figure 106 shows the hello packet format on the point-to-point network.

Figure 106 P2P IIH format

				No. of Octets
Intradomain routing protocol discriminator				1
Length indicator				1
Version/Protocol ID extension				1
ID length				1
R	R	R	PDU type	1
Version				1
Reserved				1
Maximum area address				1
Reserved/Circuit type				1
Source ID				ID length
Holding time				2
PDU length				2
Local Circuit ID				1
Variable length fields				

Instead of the priority and LAN ID fields in the LAN IIH, the P2P IIH has a Local Circuit ID field.

LSP packet format

The Link State PDUs (LSP) carries link state information. There are two types: Level-1 LSP and Level-2 LSP. The Level-2 LSP is sent by the Level-2 router, and the Level-1 LSP is sent by the Level-1 router. The level-1-2 router can sent both types of the LSPs.

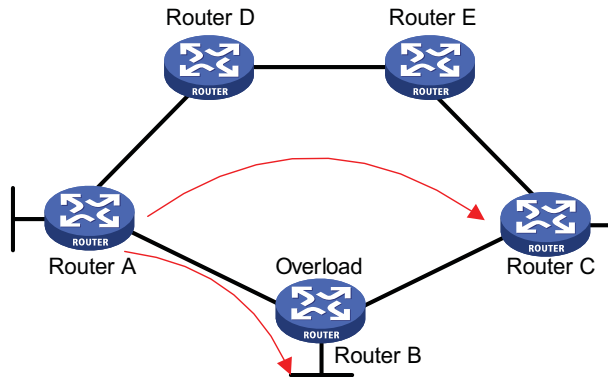
Two types of LSPs have the same format, as shown in Figure 107.

Figure 107 L1/L2 LSP format

				No. of Octets
Intradomain routing protocol discriminator				1
Length indicator				1
Version/Protocol ID extension				1
ID length				1
R	R	R	PDU type	1
Version				1
Reserved				1
Maximum area address				1
PDU length				2
Remaining lifetime				2
LSP ID				ID length+2
Sequence number				4
Checksum				2
P	ATT	OL	IS type	1
Variable length fields				

- PDU Length: Total length of the PDU in bytes.
- Remaining Lifetime: LSP remaining lifetime in seconds.
- LSP ID: Consists of the system ID, the pseudonode ID (one byte) and the LSP fragment number (one byte).
- Sequence Number: LSP sequence number.
- Checksum: LSP checksum.
- P (Partition Repair): Only related with L2 LSP, indicates whether the router supports partition repair.
- ATT (Attachment): Generated by the L1/L1 router, only related with L1 LSP, indicates that the router generating the LSP is connected with multiple areas.
- OL (LSDB Overload): Indicates that the LSDB is not complete because the router is running out of system resources. In this condition, other routers will not send packets to the overloaded router, except packets destined to the networks directly connected to the router. For example, in Figure 108, Router A uses Router B to forward its packets to Router C in normal condition. Once other routers know the OL field on Router B is set to 1, Router A will send packets to Router C via Router D and Router E, but still send to Router B packets destined to the network directly connected to Router B.

Figure 108 LSDB overload



- IS Type: Type of the router generating the LSP.

SNP format

The Sequence Number PDU (SNP) confirms the latest received LSPs. It is similar to the Acknowledge packet, but more efficient.

SNP contains Complete SNP (CSNP) and Partial SNP (PSNP), which are further divided into Level-1 CSNP, Level-2 CSNP, Level-1 PSNP and Level-2 PSNP.

CSNP covers the summary of all LSPs in the LSDB to synchronize the LSDB between neighboring routers. On broadcast networks, CSNP is sent by the DIS periodically (10s by default). On point-to-point networks, CSNP is only sent during the first adjacency establishment.

The CSNP packet format is shown in Figure 109.

Figure 109 L1/L2 CSNP format

				No. of Octets
Intradomain routing protocol discriminator				1
Length indicator				1
Version/Protocol ID extension				1
ID length				1
R	R	R	PDU type	1
Version				1
Reserved				1
Maximum area address				1
PDU length				2
Source ID				ID length+1
Start LSP ID				ID length+2
End LSP ID				ID length+2
Variable length fields				

PSNP only contains the sequence numbers of one or multiple latest received LSPs. It can acknowledge multiple LSPs at one time. When LSDBs are not synchronized, a PSNP is used to request new LSPs from neighbors.

Figure 110 shows the PSNP packet format.

Figure 110 L1/L2 PSNP format

				No. of Octets
Intradomain routing protocol discriminator				1
Length indicator				1
Version/Protocol ID extension				1
ID length				1
R	R	R	PDU type	1
Version				1
Reserved				1
Maximum area address				1
PDU length				2
Source ID				ID length+1
Variable length fields				

CLV

The variable fields of PDU are composed of multiple Code-Length-Value (CLV) triplets. Figure 111 shows the CLV format.

Figure 111 CLV format

	No. of Octets
Code	1
Length	1
Value	Length

Table 45 shows different PDUs contain different CLVs.

Table 45 CLV name and the corresponding PDU type

CLV 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, SNP
128	IP Internal Reachability Information	LSP
129	Protocols Supported	IIH, LSP
130	IP External Reachability Information	L2 LSP
131	Inter-Domain Routing Protocol Information	L2 LSP
132	IP Interface Address	IIH, LSP

Code 1 to 10 of CLV are defined in ISO 10589 (code 3 and 5 are not shown in the table), and others are defined in RFC 1195.

IS-IS Features Supported **Multiple processes**

IS-IS supports multiple processes. Multiple processes allow a IS-IS process to work in concert with a group of interfaces. This means that a router can run multiple IS-IS processes, and each process corresponds to a unique group of interfaces.

IS-IS Graceful Restart



For detailed GR information, refer to “GR Overview” on page 247.

After an IS-IS GR Restarter restarts IS-IS, it needs to complete the following two tasks to synchronize the LSDB with its neighbors.

- To obtain effective IS-IS neighbor information without changing adjacencies.
- To obtain the LSDB contents.

After the restart, the GR Restarter will send an OSPF GR signal to its neighbors to keep the adjacencies. After receiving the responses from neighbors, the GR Restarter can restore the neighbor table.

After reestablishing neighbor relationships, the GR Restarter will synchronize the LSDB and exchange routing information with all adjacent GR capable neighbors. After that, the GR Restarter will update its own routing table and forwarding table based on the new routing information and remove the stale routes. In this way, the IS-IS routing convergence is complete.

Management tag

Management tag carries the management information of the IP address prefixes and BGP community attribute. It controls the redistribution from other routing protocols.

LSP fragment extension

IS-IS advertises link state information by flooding LSPs. One LSP carries limited amount of link state information; therefore, IS-IS fragments LSPs. Each LSP fragment is uniquely identified by a combination of the System ID, Pseudonode ID (0 for a common LSP or non-zero for a Pseudonode LSP), and LSP Number (LSP fragment number) of the node or pseudo node that generated the LSP. The 1-byte LSP Number field, allowing a maximum of only 256 fragments to be generated by an IS-IS router, limits the amount of link information that the IS-IS router can advertise.

The LSP fragment extension feature allows an IS-IS router to generate more LSP fragments. Each virtual system is capable of generating 256 LSP fragments.

1 Terms

- Originating System

It is the router actually running IS-IS. After LSP fragment extension is enabled, additional virtual systems can be configured for the router. Originating system is the actual IS-IS process that originally runs.

- System ID

The system ID of the originating system.

- Additional System ID

It is the additional virtual system ID configured for the IS-IS router after LSP fragment extension is enabled. Each additional system ID can generate 256 LSP fragments. Both the additional system ID and the system ID must be unique in the entire routing domain.

- Virtual System

Virtual System is identified by the additional system ID and generates extended LSP fragments.

- Original LSP

It is the LSP generated by the originating system. The system ID in its LSP ID field is the system ID of the originating system.

- Extended LSP

It is the LSP generated by a virtual system. The system ID in its LSP ID field is the virtual system ID.

After additional system IDs are configured, an IS-IS router can advertise more link state information in extended LSP fragments. Each virtual system can be considered as a virtual router. An extended LSP fragment is advertised by a virtual system identified by additional system ID.

2 Operation modes

The LSP fragment extension feature operates in two modes on an IS-IS router:

- Mode-1: It applies to a network where some routers do not support LSP fragment extension. In this mode, adjacency is formed between the originating system and each virtual system, with the link cost from the originating system to each virtual system as 0. Thus, each virtual system acts as a router connected to the originating system in the network, but the virtual system is reachable through the originating system only. Therefore, the IS-IS routers not supporting LSP fragment extension can operate normally without modifying the extended LSP fragments received, but some limitation is imposed on the link state information in the extended LSP fragments advertised by the virtual systems.
- Mode-2: This mode is recommended in a network where all the routers support LSP fragment extension. In this mode, all the IS-IS routers in the network know which originating system the LSPs generated by the virtual systems belong to; therefore no limitation is imposed on the link state information of the extended LSP fragments advertised by the virtual systems.

The operation mode of LSP fragment extension is configured based on area and routing level. Mode-1 allows the routers supporting and not supporting LSP fragment extension to interoperate with each other, but it restricts the link state information in the extended fragments. Mode-2 does not restrict the link state

information in the extended fragments. Mode-2 is recommended in a network where all the routers that are in the same area and at the same routing level support LSP fragment extension.

Dynamic host name mapping mechanism

The dynamic host name mapping mechanism provides the mapping between the host names and the system IDs for the IS-IS routers. The dynamic host name information is announced in the dynamic host name CLV of an LSP.

This mechanism also provides the mapping between a host name and the DIS of a broadcast network, which is announced in a dynamic host name TLV of a pseudonode LSP.

A host name is intuitively easier to remember than a system ID. After enabling this feature on the router, you can see the host names instead of system IDs using the **display** command.

Protocols and Standards

- ISO 10589 ISO IS-IS Routing Protocol
- ISO 9542 ES-IS Routing Protocol
- ISO 8348/Ad2 Network Services Access Points
- RFC 1195 - Use of OSI IS-IS for Routing in TCP/IP and Dual Environments
- 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 3358 - Optional Checksums in ISIS
- RFC 3373 - Three-Way Handshake for IS-IS Point-to-Point Adjacencies
- RFC 3567 - Intermediate System to Intermediate System (IS-IS) Cryptographic Authentication
- RFC 3719 - Recommendations for Interoperable Networks using IS-IS
- RFC 3786 - Extending the Number of IS-IS LSP Fragments Beyond the 256 Limit
- RFC 3787 - Recommendations for Interoperable IP Networks using IS-IS
- RFC 3847 - Restart signaling for IS-IS

IS-IS Configuration Task List

Complete the following tasks to configure IS-IS:

Task	Remarks
"Configuring IS-IS Basic Functions" on page 341	Required

Task	Remarks	
"Configuring IS-IS Routing Information Control" on page 342	"Specifying a Priority for IS-IS" on page 342	Optional
	"Configuring IS-IS Link Cost" on page 343	Required
	"Configuring the Maximum Number of Equal Cost Routes" on page 344	Optional
	"Configuring IS-IS Route Summarization" on page 344	Optional
	"Advertising a Default Route" on page 345	Optional
	"Configuring Inbound Route Filtering" on page 345	Optional
	"Configuring Route Redistribution" on page 345	Optional
"Tuning and Optimizing IS-IS Network" on page 346	"Configuring IS-IS Route Leaking" on page 346	Optional
	"Configuring a DIS Priority for an Interface" on page 346	Optional
	"Configuring IS-IS Timers" on page 347	Optional
	"Disabling an Interface from Sending/Receiving IS-IS Hello Packets" on page 347	Optional
	"Configuring LSP Parameters" on page 348	Optional
	"Configuring SPF Parameters" on page 349	Optional
	"Configuring Dynamic Host Name Mapping" on page 349	Optional
	"Configuring IS-IS Authentication" on page 350	Optional
	"Configuring LSDB Overload Tag" on page 351	Optional
	"Logging the Adjacency Changes" on page 351	Optional
"Enabling an Interface to Send Small Hello Packets" on page 351	Optional	
"Enabling SNMP Trap" on page 352	Optional	
"Configuring IS-IS GR" on page 352	Optional	

Configuring IS-IS Basic Functions

Configuration Prerequisites Before the task, configure an IP address for each interface, making all adjacent nodes reachable to each other at the network layer.

Configuration Procedure Follow these steps to configure IS-IS basic functions:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enable IS-IS routing process and enter its view	isis [<i>process-id</i>]	Required Not enabled by default

To do...	Use the command...	Remarks
Assign a network entity title (NET)	network-entity <i>net</i>	Required Not assigned by default
Specify a router type	is-level { level-1 level-1-2 level-2 }	Optional The default type is level-1-2.
Return to system view	quit	--
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--
Enable an IS-IS process on the interface	isis enable [<i>process-id</i>]	Required Disabled by default
Specify network type for the interface as P2P	isis circuit-type p2p	Optional By default, the network type of an interface depends on the physical media. The network type of a VLAN interface is broadcast.
Specify the adjacency type for the interface	isis circuit-level [level-1 level-1-2 level-2]	Optional The default type is level-1-2.



*If a router's type is configured as Level-1 or Level-2, the type of interfaces must be the same, which cannot be changed using the **isis circuit-level** command. However, an interface's type can be changed with this command when the router's type is Level-1-2 for the establishment of a specific level adjacency.*

Configuring IS-IS Routing Information Control

Configuration Prerequisites

Before the configuration, accomplish the following tasks first:

- Configure an IP address on each interface, and make sure all nodes are reachable.
- Configure basic IS-IS functions

Specifying a Priority for IS-IS

A router can run multiple routing protocols. When a route to the same destination is learned by multiple routing protocols, the one with the highest protocol priority wins. You can reference a routing policy to specify a priority for specific routes. For information about routing policy, refer to "Routing Policy Configuration" on page 415.

Follow these steps to configure the IS-IS protocol priority.

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter IS-IS view	isis [<i>process-id</i>]	--
Specify a priority for IS-IS	preference { route-policy <i>route-policy-name</i> <i>preference</i> } *	Optional 15 by default

Configuring IS-IS Link Cost

There are three ways to configure the interface link cost, in descending order of interface costs:

- Interface cost: Assign a link cost for a single interface.
- Global cost: Assign a link cost for all interfaces.
- Automatically calculated cost: Calculate the link cost based on the bandwidth of an interface.

Interface cost defaults to 10.

Configure an IS-IS cost for an interface

Follow these steps to configure an interface's cost:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter IS-IS view	isis [<i>process-id</i>]	--
Specify a cost style	cost-style { narrow wide wide-compatible { compatible narrow-compatible } [relax-spf-limit] }	Optional narrow by default
Return to system view	quit	--
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	Required
Specify a cost for the interface	isis cost <i>value</i> [level-1 level-2]	Optional Not specified by default

Configure a global IS-IS cost

Follow these steps to configure global IS-IS cost:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter IS-IS view	isis [<i>process-id</i>]	-
Specify an IS-IS cost style	cost-style { narrow wide wide-compatible { compatible narrow-compatible } [relax-spf-limit] }	Optional Defaulted as narrow.
Specify a global IS-IS cost	circuit-cost <i>value</i> [level-1 level-2]	Required Not specified by default.

Enable automatic IS-IS cost calculation

Follow these steps to enable automatic IS-IS cost calculation:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter IS-IS view	isis [<i>process-id</i>]	-

To do...	Use the command...	Remarks
Specify an IS-IS cost style	cost-style { narrow wide wide-compatible { compatible narrow-compatible } [relax-spf-limit] }	Optional narrow by default
Configure a bandwidth reference value for automatic IS-IS cost calculation	bandwidth-reference <i>value</i>	Optional 100 Mbps by default
Enable automatic IS-IS cost calculation	auto-cost enable	Required Disabled by default.



In the case no interface cost is specified in interface view or system view and automatic cost calculation is enabled

- When the cost style is **wide** or **wide-compatible**, IS-IS automatically calculates the interface cost based on the interface bandwidth, using the formula: $\text{interface cost} = \text{bandwidth reference value} / \text{interface bandwidth}$, and the maximum calculated cost is 16777214.
- When the cost style is **narrow**, **narrow-compatible**, or **compatible**, if the interface is a loopback interface, the cost value is 0; otherwise, the cost value is automatically calculated as follows: if the interface bandwidth is in the range of 1 M to 10 M, the interface cost is 60; if the interface bandwidth is in the range of 11 M to 100 M, the interface cost is 50; if the interface bandwidth is in the range of 101 M to 155 M, the interface cost is 40; if the interface bandwidth is in the range of 156 M to 622 M, the interface cost is 30; if the interface bandwidth is in the range of 623 M to 2500 M, the interface cost is 20, and the default interface cost of 10 is used for any other bandwidths.

Configuring the Maximum Number of Equal Cost Routes

If there are more than one equal cost routes to the same destination, the traffic can be load balanced to enhance efficiency.

Follow these steps to configure the maximum number of equal cost routes:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter IS-IS view	isis [<i>process-id</i>]	--
Specify the maximum number of equal cost routes for load balancing	maximum load-balancing <i>number</i>	Optional The default number is 4.

Configuring IS-IS Route Summarization

This task is to configure a summary route, so routes falling into the network range of the summary route are summarized with one route for advertisement. Doing so can reduce the size of routing tables, as well as the LSP and LSDB generated by the router itself. Both IS-IS and redistributed routes can be summarized.

Follow these steps to configure route summarization:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Enter IS-IS view	isis [<i>process-id</i>]	--
Configure IS-IS route summarization	summary <i>ip-address</i> { <i>mask</i> <i>mask-length</i> } [avoid-feedback generate_null0_route tag <i>tag</i> [level-1 level-1-2 level-2]] *	Required Not configured by default



The cost of the summary route is the lowest cost among those summarized routes.

Advertising a Default Route

Follow these steps to advertise a default route:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter IS-IS view	isis [<i>process-id</i>]	-
Advertise a default route	default-route-advertise [route-policy <i>route-policy-name</i>] [level-1 level-2 level-1-2]	Optional Level-2 router generates a default route by default.



The default route is only advertised to routers at the same level. You can use a routing policy to generate the default route only when a local routing entry is matched by the policy.

Configuring Inbound Route Filtering

Follow these steps to configure inbound route filtering:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter IS-IS view	isis [<i>process-id</i>]	--
Configure inbound route filtering	filter-policy { <i>acl-number</i> ip-prefix <i>ip-prefix-name</i> route-policy <i>route-policy-name</i> } import	Required Not configured by default

Configuring Route Redistribution

Follow these steps to configure IS-IS route redistribution from other routing protocols:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter IS-IS view	isis [<i>process-id</i>]	--
Redistribute routes from another routing protocol	import-route { isis [<i>process-id</i>] ospf [<i>process-id</i>] rip [<i>process-id</i>] bgp [allow-ibgp] direct static } [cost <i>cost</i> cost-type { external internal }] [level-1 level-1-2 level-2] route-policy <i>route-policy-name</i> tag <i>tag</i>] *	Required No route is redistributed by default. If no level is specified, routes are redistributed into the Level-2 routing table by default.

To do...	Use the command...	Remarks
Configure a filtering policy to filter redistributed routes	filter-policy { <i>acl-number</i> ip-prefix <i>ip-prefix-name</i> route-policy <i>route-policy-name</i> } export [isis <i>process-id</i> ospf <i>process-id</i> rip <i>process-id</i> bgp direct static]	Optional Not configured by default

Configuring IS-IS Route Leaking

With this feature enabled, the Level-1-2 router can advertise both Level-1 and Level-2 area routing information to the Level-1 router.

Follow these steps to configure IS-IS route leaking:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter IS-IS view	isis [<i>process-id</i>]	--
Enable IS-IS route leaking	import-route isis level-2 into level-1 [filter-policy { <i>acl-number</i> ip-prefix <i>ip-prefix-name</i> route-policy <i>route-policy-name</i> } tag <i>tag</i>] *	Required Disabled by default



- If a filter policy is specified, only routes passing it can be advertised into Level-1 area.
- You can specify a routing policy in the **import-route isis level-2 into level-1** command to filter routes from Level-2 to Level-1. Other routing policies specified for route reception and redistribution does not affect the route leaking.

Tuning and Optimizing IS-IS Network

Configuration Prerequisites

Before the configuration, accomplish the following tasks first:

- Configure an IP address on each interface, and make sure all nodes are reachable.
- Configure basic IS-IS functions

Configuring a DIS Priority for an Interface

On an IS-IS broadcast network, a router should be selected as the DIS at a specific level, Level-1 or Level-2. You can specify a DIS priority at a level for an interface. The bigger the interface's priority value, the more likelihood it becomes the DIS.

Follow these steps to configure a DIS priority for an interface:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--
Specify a DIS priority for the interface	isis dis-priority <i>value</i> [level-1 level-2]	Optional 64 by default



If multiple routers in the broadcast network have the same highest DIS priority, the router with the highest MAC address becomes the DIS. This rule applies even all routers' DIS priority is 0.

Configuring IS-IS Timers

Follow these steps to configure the IS-IS timers:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--
Specify the interval between hello packets	isis timer hello <i>seconds</i> [level-1 level-2]	Optional 10 seconds by default
Specify the number of hello packets; within the time for receiving the specified hello packets, if no hello packets are received on the interface, the neighbor is considered dead.	isis timer holding-multiplier <i>value</i> [level-1 level-2]	Optional 3 by default
Specify the interval for sending CSNP packets	isis timer csnp <i>seconds</i> [level-1 level-2]	Optional 10 seconds by default
Specify the interval for sending LSP packets	isis timer lsp <i>time</i> [count <i>count</i>]	Optional 33 milliseconds by default
Specify the LSP retransmission interval on the point-to-point link	isis timer retransmit <i>seconds</i>	Optional 5 seconds by default



- On the broadcast link, you can specify different intervals for Level-1 and Level-2 hello packets; if no level is specified, the interval applies to both Level-1 and Level-2 hello packets, but only takes effect on the level of the current process; if a level is specified, it applies to hello packets at this level. The point-to-point link does not distinguish between Level-1 and Level-2 hello packets, so you need not specify a level.
- Hello packets are used to establish and maintain neighbor relationships. If no hello packets are received from a neighbor within the time for receiving the specified hello packets, the neighbor is considered dead.
- CSNPs are sent by the DIS on a broadcast network for LSDB synchronization. If no level is included, the specified CSNP interval applies to both Level-1 and Level-2 of the current IS-IS process. If a level is specified, it applies to the level.
- On a point-to-point link, if there is no response to a LSP sent by the local router within the specified retransmission interval, the LSP is considered lost, and the same LSP will be retransmitted. On broadcast links, responses to the sent LSPs are not required.
- The interval between hello packets sent by the DIS is 1/3 the hello interval set by the **isis timer hello** command.

Disabling an Interface from Sending/Receiving IS-IS Hello Packets

Follow these steps to disable an interface from sending hello packets:

To do...	Use the command...	Remarks
Enter system view	system-view	--

To do...	Use the command...	Remarks
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--
Disable the interface from sending and receiving hello packets	isis silent	Required Not disabled by default

Configuring LSP Parameters

An IS-IS router periodically advertises all the local LSPs to maintain the LSP synchronization in the entire area.

A LSP is given an aging time when generated by the router. When the LSP is received by another router, its aging time begins to decrease. If the receiving router does not get the update for the LSP within the aging time, the LSP will be deleted from the LSDB.

The router will discard a LSP with incorrect checksum. You can configure the router to ignore the incorrect checksum, which means a LSP will be processed even with an incorrect LSP checksum.

On the NBMA network, the router will flood a new LSP received from an interface to other interfaces. This can cause the LSP reflooding on the high connectivity networks. To avoid this problem, you can make a mesh group of interfaces. The interface in this group will only flood the new LSP to interfaces outside the mesh group.

Follow these steps to configure the LSP parameters:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter IS-IS view	isis [<i>process-id</i>]	--
Specify a LSP refresh interval	timer lsp-refresh <i>seconds</i>	Optional 900 seconds by default
Specify the maximum LSP aging time	timer lsp-max-age <i>seconds</i>	Optional 1200 seconds by default
Specify LSP generation interval	timer lsp-generation <i>maximum-interval</i> [<i>initial-interval</i> [<i>incremental-interval</i>]] [level-1 level-2]	Optional 2 seconds by default
Enable the LSP flash flooding function	flash-flood [flood-count <i>flood-count</i> max-timer-interval <i>flood-interval</i> [level-1 level-2]] *	Optional Not enabled by default
Specify the maximum size of the originated Level-1 or Level-2 LSP	lsp-length originate <i>size</i> [level-1 level-2]	Optional Both are 1497 bytes by default
Specify the maximum size of the received Level-1 or Level-2 LSP	lsp-length receive <i>size</i>	Optional Both are 1497 bytes by default

To do...	Use the command...	Remarks
Enable LSP fragment extension	lsp-fragments-extend [level-1 level-2 level-1-2] [mode-1 mode-2]	Optional Disabled by default
Create a virtual system	virtual-system <i>virtual-system-id</i>	Optional Not created by default
Return to system view	quit	--
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--
Add the interface to a mesh group	isis mesh-group [<i>mesh-group-number</i>] mesh-blocked]	Optional Not added by default If the mesh-blocked keyword is included, the interface is blocked from flooding LSPs. It can send an LSP only after receiving a request.



Note the following when enabling LSP fragment extension

- After LSP fragment extension is enabled in an IS-IS process, the MTUs of all the interfaces with this IS-IS process enabled must not be less than 512; otherwise, LSP fragment extension will not take effect.
- At least one virtual system needs to be configured for the router to generate extended LSP fragments.

Configuring SPF Parameters

When the LSDB changes in an IS-IS network, a routing calculation starts. If the changes happen frequently, it will take a lot of system resources. You can set the interval for SPF calculation for efficiency consideration.

The SPF calculation may occupy the CPU for a long time when the routing entries are too many. You can split the SPF calculation time into multiple durations with a default interval of 10s in between.

Follow these steps to configure the SPF parameters:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter IS-IS view	isis [<i>process-id</i>]	--
Configure the SPF calculation intervals	timer spf <i>maximum-interval</i> [<i>minimum-interval</i> [<i>incremental-interval</i>]]	Optional The default SPF calculation interval is 10 seconds.
Specify the SPF calculation duration	spf-slice-size <i>duration-time</i>	Optional 10 milliseconds by default

Configuring Dynamic Host Name Mapping

Follow these steps to configure the dynamic host name mapping:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter IS-IS view	isis [<i>process-id</i>]	--

To do...	Use the command...	Remarks
Assign a local host name	is-name <i>sys-name</i>	Required No name is assigned by default. This command also enables the mapping between the local system ID and host name
Assign a remote host name and create a mapping between the host name and a system ID	is-name map <i>sys-id map-sys-name</i>	Optional One system ID only maps to one name. No name is assigned by default
Return to system view	quit	--
Enter interface view	interface <i>interface-type interface-number</i>	--
Assign a DIS name for the local network	isis dis-name <i>symbolic-name</i>	Optional Not assigned by default This command is only applicable on the router with dynamic host name mapping enabled. It is invalid on point-to-point links.



The local host name on the local IS overwrites the remote host name on the remote IS.

Configuring IS-IS Authentication

For area authentication, the area authentication password is encapsulated into the Level-1 LSP, CSNP, and PSNP packets. On area authentication enabled routers in the same area, the authentication mode and password must be same.

For routing domain authentication, the domain authentication password is encapsulated into the Level-2 LSP, CSNP, and PSNP packets. The domain authentication enabled Level-2 routers in the backbone must adopt the same authentication mode and share the same password.

The authentication configured on an interface applies to the hello packet in order to authenticate neighbors. All interfaces within a network must share the same authentication password at the same level.

Follow these steps to configure the authentication function:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter IS-IS view	isis [<i>process-id</i>]	--
Specify the area authentication mode	area-authentication-mode { simple md5 } <i>password</i> [ip osi]	Required No authentication is enabled for Level-1 routing information, and no password is specified by default.

To do...	Use the command...	Remarks
Specify the routing domain authentication mode	domain-authentication-mode { simple md5 } <i>password</i> [ip osi]	Required No authentication is enabled for Level-2 routing information, and no password is specified by default.
Return to system view	quit	--
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--
Specify the authentication mode and password	isis authentication-mode { simple md5 } <i>password</i> [level-1 level-2] [ip osi]	Optional No authentication and password are available by default.



The **level-1** and **level-2** keywords in the **isis authentication-mode** command are only supported on a VLAN interface of a switch, and the interface must be configured with the **isis enable** command first.

Configuring LSDB Overload Tag

When the overload tag is set on a router, other routers will not send packets to the router except for the packets destined to the network directly connected to the router.

The overload tag can be used for troubleshooting as well. You can temporarily isolate a router from the IS-IS network by setting the overload tag.

Follow these steps to configure the LSDB overload tag:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter IS-IS view	isis [<i>process-id</i>]	--
Configure the overload tag	set-overload [on-startup [[start-from-nbr <i>system-id</i> [<i>timeout1</i> [<i>nbr-timeout</i>]]]] [<i>timeout2</i>] [allow { interlevel external } *]	Required Not configured by default

Logging the Adjacency Changes

Follow these steps to configure this task:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter IS-IS view	isis [<i>process-id</i>]	--
Enable to log the adjacency changes	log-peer-change	Required Enabled by default



With this feature enabled, the state information of the adjacency is displayed on the configuration terminal.

Enabling an Interface to Send Small Hello Packets

Follow these steps to enable an interface to send small hello packets (without the padding field):

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--
Enable the interface to send small hello packets that have no padding field	isis small-hello	Required Standard hello packets are sent by default.

Enabling SNMP Trap Follow these steps to enable IS-IS trap:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter IS-IS view	isis [<i>process-id</i>]	--
Enable SNMP Trap	is-snmp-traps enable	Required Enabled by default

Configuring IS-IS GR

An ISIS restart may cause the termination of the adjacencies between a restarting router and its neighbors, resulting in a transient network disconnection.

IS-IS Graceful Restart can help to solve this problem by notifying its neighbors its restarting state to allow them to reestablish the adjacency without removing it. The IS-IS Graceful Restart provides the following features:

- When restarting ISIS, a Graceful Restart capable device will resend connection requests to its neighbors instead of terminating their neighboring relationships.
- Graceful Restart minimizes network disruption caused by LSDB synchronization before LSP packets generation.
- When a router starts for the first time, it sets the overload bit in LSP packets before LSDB synchronization is complete, which ensures no routing loop is created.

The Graceful Restart interval on a router is used as the holdtime in the IS-IS Hello PDUs so that its neighbors can maintain the adjacencies within the interval after the router restarts.

By setting the SA (Suppress-Advertisement) bit in the hello PDUs sent by the GR Restarter, its neighbors will not advertise adjacencies within the specified period until the completion of LSDB synchronization between the GR Restarter and its neighbors. This feature helps to effectively avoid blackhole routes due to the sending or receiving of LSPs across the restart.



A device can act as both the GR Restarter and GR Helper at the same time.

Follow these steps to configure GR on the GR Restarter and GR Helper respectively:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Enable IS-IS, and enter IS-IS view	isis [<i>process-id</i>]	Required Disabled by default
Enable the GR capability for IS-IS	graceful-restart	Required Disabled by default
Set the Graceful Restart interval	graceful-restart interval <i>timer</i>	Required 300 seconds by default
Configure to set the SA bit during restart	graceful-restart suppress-sa	Optional By default, the SA bit is not set.

Displaying and Maintaining IS-IS

To do...	Use the command...	Remarks
Display brief IS-IS information	display isis brief [<i>process-id</i>]	Available in any view
Display information about IS-IS enabled interfaces	display isis interface [verbose] [<i>process-id</i>]	Available in any view
Display IS-IS license information	display isis license	Available in any view
Display IS-IS LSDB information	display isis lsdb [[I1 I2 level-1 level-2]] [lsp-id <i>LSPID</i> lsp-name <i>lspname</i>] local verbose] * [<i>process-id</i>]	Available in any view
Display IS-IS mesh group information	display isis mesh-group [<i>process-id</i>]	Available in any view
Display the host-name-to-system-ID mapping table	display isis name-table [<i>process-id</i>]	Available in any view
Display IS-IS neighbor information	display isis peer [verbose] [<i>process-id</i>]	Available in any view
Display IS-IS routing information	display isis route [ipv4] [[level-1 level-2]] verbose] * [<i>process-id</i>]	Available in any view
Display SPF calculation log information	display isis spf-log [<i>process-id</i>]	Available in any view
Display the statistics about an IS-IS process	display isis statistics [level-1 level-2 level-1-2] [<i>process-id</i>]	Available in any view
Display the IS-IS Graceful Restart state	display isis graceful-restart status [level-1 level-2] [<i>process-id</i>]	Available in any view
Clear the data structure information of an IS-IS process	reset isis all [<i>process-id</i>]	Available in user view
Clear the data structure information of an IS-IS neighbor	reset isis peer <i>system-id</i> [<i>process-id</i>]	Available in user view

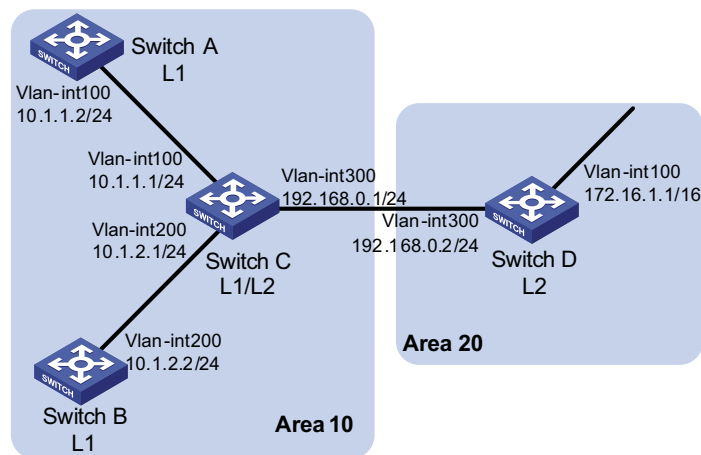
IS-IS Configuration Example

IS-IS Basic Configuration Network requirements

As shown in Figure 112, Switch A, B, C and Switch D reside in an IS-IS AS. Switch A and B are Level-1 switches, Switch D is a Level-2 switch and Switch C is a Level-1-2 switch. Switch A, B and C are in area 10, while Switch D is in area 20.

Network diagram

Figure 112 Network diagram for IS-IS basic configuration



Configuration procedure

- 1 Configure IP addresses for interfaces (omitted)
- 2 Configure IS-IS

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] is-level level-1
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis enable 1
[SwitchA-Vlan-interface100] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] is-level level-1
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis enable 1
[SwitchB-Vlan-interface200] quit
```

Configure Switch C.

```

<SwitchC> system-view
[SwitchC] isis 1
[SwitchC-isis-1] network-entity 10.0000.0000.0003.00
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] isis enable 1
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis enable 1
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] isis enable 1
[SwitchC-Vlan-interface300] quit

```

Configure Switch D.

```

<SwitchD> system-view
[SwitchD] isis 1
[SwitchD-isis-1] is-level level-2
[SwitchD-isis-1] network-entity 20.0000.0000.0004.00
[SwitchD-isis-1] quit
[SwitchD] interface vlan-interface 100
[SwitchD-Vlan-interface100] isis enable 1
[SwitchD-Vlan-interface100] quit
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] isis enable 1
[SwitchD-Vlan-interface300] quit

```

3 Verify the configuration

Display the IS-IS LSDB of each switch to check the LSP integrity.

```
[SwitchA] display isis lsdb
```

```

Database information for ISIS(1)
-----

Level-1 Link State Database

LSPID                Seq Num      Checksum    Holdtime    Length  ATT/P/OL
-----
0000.0000.0001.00-00* 0x00000004  0xdf5e     1096       68     0/0/0
0000.0000.0002.00-00 0x00000004  0xee4d     1102       68     0/0/0
0000.0000.0002.01-00 0x00000001  0xdaaf     1102       55     0/0/0
0000.0000.0003.00-00 0x00000009  0xcaa3     1161       111    1/0/0
0000.0000.0003.01-00 0x00000001  0xadda     1112       55     0/0/0

```

*-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload

```
[SwitchB] display isis lsdb
```

```

Database information for ISIS(1)
-----

Level-1 Link State Database

LSPID                Seq Num      Checksum    Holdtime    Length  ATT/P/OL
-----
0000.0000.0001.00-00 0x00000006  0xdb60     988        68     0/0/0
0000.0000.0002.00-00* 0x00000008  0xe651     1189       68     0/0/0
0000.0000.0002.01-00* 0x00000005  0xd2b3     1188       55     0/0/0
0000.0000.0003.00-00 0x00000014  0x194a     1190       111    1/0/0
0000.0000.0003.01-00 0x00000002  0xabdb     995        55     0/0/0

```

```

*-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload

[SwitchC] display isis lsdb

```

```

Database information for ISIS(1)
-----

Level-1 Link State Database

LSPID                Seq Num      Checksum     Holdtime     Length  ATT/P/OL
-----
0000.0000.0001.00-00 0x00000006  0xdb60      847          68      0/0/0
0000.0000.0002.00-00 0x00000008  0xe651      1053         68      0/0/0
0000.0000.0002.01-00 0x00000005  0xd2b3      1052         55      0/0/0
0000.0000.0003.00-00* 0x00000014  0x194a      1051         111     1/0/0
0000.0000.0003.01-00* 0x00000002  0xabdb      854          55      0/0/0

```

```

*-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload

```

```

Level-2 Link State Database

LSPID                Seq Num      Checksum     Holdtime     Length  ATT/P/OL
-----
0000.0000.0003.00-00* 0x00000012  0xc93c      842          100     0/0/0
0000.0000.0004.00-00 0x00000026  0x331       1173         84      0/0/0
0000.0000.0004.01-00 0x00000001  0xee95      668          55      0/0/0

```

```

*-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload

```

```

[SwitchD] display isis lsdb

Database information for ISIS(1)
-----

Level-2 Link State Database

LSPID                Seq Num      Checksum     Holdtime     Length  ATT/P/OL
-----
0000.0000.0003.00-00 0x00000013  0xc73d      1003         100     0/0/0
0000.0000.0004.00-00* 0x0000003c  0xd647      1194         84      0/0/0
0000.0000.0004.01-00* 0x00000002  0xec96      1007         55      0/0/0

```

```

*-Self LSP, +-Self LSP(Extended), ATT-Attached, P-Partition, OL-Overload

```

Display the IS-IS routing information of each switch. Level-1 switches should have a default route with the next hop being the Level-1-2 switch. The Level-2 switch should have both routing information of Level-1 and Level-2.

```

[SwitchA] display isis route

Route information for ISIS(1)
-----

ISIS(1) IPv4 Level-1 Forwarding Table
-----

IPV4 Destination    IntCost    ExtCost    ExitInterface    NextHop      Flags
-----
10.1.1.0/24         10         NULL      Vlan100          Direct       D/L/-
10.1.2.0/24         20         NULL      Vlan100          10.1.1.1    R/-/-
192.168.0.0/24     20         NULL      Vlan100          10.1.1.1    R/-/-
0.0.0.0/0           10         NULL      Vlan100          10.1.1.1    R/-/-

```

```

Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set

```

```

[SwitchC] display isis route

```


Route information for ISIS(1)

ISIS(1) IPv4 Level-1 Forwarding Table

IPv4 Destination	IntCost	ExtCost	ExitInterface	NextHop	Flags
192.168.0.0/24	10	NULL	Vlan300	Direct	D/L/-
10.1.1.0/24	10	NULL	Vlan100	Direct	D/L/-
10.1.2.0/24	10	NULL	Vlan200	Direct	D/L/-

Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set

ISIS(1) IPv4 Level-2 Forwarding Table

IPv4 Destination	IntCost	ExtCost	ExitInterface	NextHop	Flags
192.168.0.0/24	10	NULL	Vlan300	Direct	D/L/-
10.1.1.0/24	10	NULL	Vlan100	Direct	D/L/-
10.1.2.0/24	10	NULL	Vlan200	Direct	D/L/-
172.16.0.0/16	20	NULL	Vlan300	192.168.0.2	R/-/-

Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set

[SwitchD] display isis route

Route information for ISIS(1)

ISIS(1) IPv4 Level-2 Forwarding Table

IPv4 Destination	IntCost	ExtCost	ExitInterface	NextHop	Flags
192.168.0.0/24	10	NULL	Vlan300	Direct	D/L/-
10.1.1.0/24	20	NULL	Vlan300	192.168.0.1	R/-/-
10.1.2.0/24	20	NULL	Vlan300	192.168.0.1	R/-/-
172.16.0.0/16	10	NULL	Vlan100	Direct	D/L/-

Flags: D-Direct, R-Added to RM, L-Advertised in LSPs, U-Up/Down Bit Set

DIS Selection Configuration

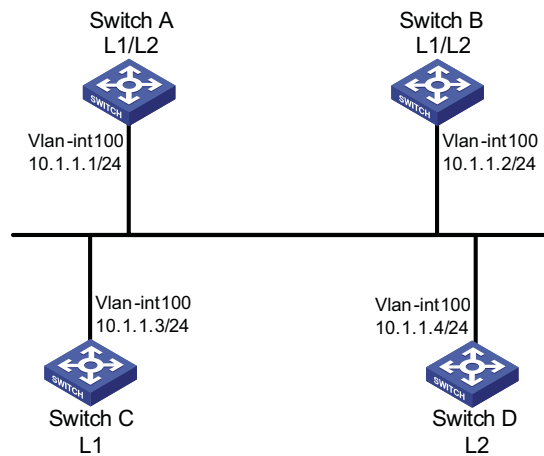
Network requirements

As shown in Figure 113, Switch A, B, C and Switch D reside in IS-IS area 10 on a broadcast network (Ethernet). Switch A and Switch B are Level-1-2 switches, Switch C is a Level-1 switch, and Switch D is a Level-2 switch.

Change the DIS priority of Switch A to make it selected as the Level-1-2 DIS router.

Network diagram

Figure 113 Network diagram for DIS selection



Configuration procedure

- 1 Configure an IP address for each interface (omitted)
- 2 Enable IS-IS

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis enable 1
[SwitchA-Vlan-interface100] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] isis enable 1
[SwitchB-Vlan-interface100] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis 1
[SwitchC-isis-1] network-entity 10.0000.0000.0003.00
[SwitchC-isis-1] is-level level-1
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] isis enable 1
[SwitchC-Vlan-interface100] quit
```

Configure Switch D.

```

<SwitchD> system-view
[SwitchD] isis 1
[SwitchD-isis-1] network-entity 10.0000.0000.0004.00
[SwitchD-isis-1] is-level level-2
[SwitchD-isis-1] quit
[SwitchD] interface vlan-interface 100
[SwitchD-Vlan-interface100] isis enable 1
[SwitchD-Vlan-interface100] quit

```

Display information about IS-IS neighbors of Switch A.

```

[SwitchA] display isis peer

Peer information for ISIS(1)
-----
System Id: 0000.0000.0002
Interface: Vlan-interface100      Circuit Id: 0000.0000.0003.01
State: Up      HoldTime: 21s      Type: L1(L1L2)      PRI: 64

System Id: 0000.0000.0003
Interface: Vlan-interface100      Circuit Id: 0000.0000.0003.01
State: Up      HoldTime: 27s      Type: L1              PRI: 64

System Id: 0000.0000.0002
Interface: Vlan-interface100      Circuit Id: 0000.0000.0004.01
State: Up      HoldTime: 28s      Type: L2(L1L2)      PRI: 64

System Id: 0000.0000.0004
Interface: Vlan-interface100      Circuit Id: 0000.0000.0004.01
State: Up      HoldTime: 30s      Type: L2              PRI: 64

```

Display information about IS-IS interfaces of Switch A.

```

[SwitchA] display isis interface

Interface information for ISIS(1)
-----
Interface: Vlan-interface100
Id      IPV4.State  IPV6.State  MTU   Type  DIS
001      Up          Down        1497  L1/L2 No/No

```

Display information about IS-IS interfaces of Switch C.

```

[SwitchC] display isis interface

Interface information for ISIS(1)
-----
Interface: Vlan-interface100
Id      IPV4.State  IPV6.State  MTU   Type  DIS
001      Up          Down        1497  L1/L2 Yes/No

```

Display information about IS-IS interfaces of Switch D.

```

[SwitchD] display isis interface

Interface information for ISIS(1)
-----
Interface: Vlan-interface100

```

Id	IPV4.State	IPV6.State	MTU	Type	DIS
001	Up	Down	1497	L1/L2	No/Yes



By using the default DIS priority, Switch C is the Level-1 DIS, and Switch D is the Level-2 DIS. The pseudo nodes of Level-1 and Level-2 are 0000.0000.0003.01 and 0000.0000.0004.01 respectively.

3 Configure the DIS priority of Switch A.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis dis-priority 100
[SwitchA-Vlan-interface100] quit
```

Display IS-IS neighbors of Switch A.

```
[SwitchA] display isis peer
```

```

Peer information for ISIS(1)
-----
System Id: 0000.0000.0002
Interface: Vlan-interface100      Circuit Id: 0000.0000.0001.01
State: Up      HoldTime: 21s      Type: L1(L1L2)      PRI: 64

System Id: 0000.0000.0003
Interface: Vlan-interface100      Circuit Id: 0000.0000.0001.01
State: Up      HoldTime: 27s      Type: L1              PRI: 64

System Id: 0000.0000.0002
Interface: Vlan-interface100      Circuit Id: 0000.0000.0001.01
State: Up      HoldTime: 28s      Type: L2(L1L2)      PRI: 64

System Id: 0000.0000.0004
Interface: Vlan-interface100      Circuit Id: 0000.0000.0001.01
State: Up      HoldTime: 30s      Type: L2              PRI: 64

```

Display information about IS-IS interfaces of Switch A.

```
[SwitchA] display isis interface
```

```

Interface information for ISIS(1)
-----
Interface: Vlan-interface100
Id      IPV4.State  IPV6.State  MTU  Type  DIS
001     Up         Down       1497 L1/L2 Yes/Yes

```



After the DIS priority configuration, Switch A becomes the Level-1-2 DIS, and the pseudonode is 0000.0000.0001.01.

Display information about IS-IS neighbors and interfaces of Switch C.

```
[SwitchC] display isis peer
```

```

Peer information for ISIS(1)
-----
System Id: 0000.0000.0002
Interface: Vlan-interface100      Circuit Id: 0000.0000.0001.01

```

```

State: Up      HoldTime: 25s      Type: L1      PRI: 64

System Id: 0000.0000.0001
Interface: Vlan-interface100      Circuit Id: 0000.0000.0001.01
State: Up      HoldTime: 7s      Type: L1      PRI: 100

```

```
[SwitchC] display isis interface
```

```

                          Interface information for ISIS(1)
                          -----
Interface: Vlan-interface100
Id      IPV4.State      IPV6.State      MTU      Type      DIS
001     Up              Down            1497     L1/L2     No/No

```

```
# Display information about IS-IS neighbors and interfaces of Switch D.
```

```
[SwitchD] display isis peer
```

```

                          Peer information for ISIS(1)
                          -----
System Id: 0000.0000.0001
Interface: Vlan-interface100      Circuit Id: 0000.0000.0001.01
State: Up      HoldTime: 9s      Type: L2      PRI: 100

System Id: 0000.0000.0002
Interface: Vlan-interface100      Circuit Id: 0000.0000.0001.01
State: Up      HoldTime: 28s     Type: L2      PRI: 64

```

```
[SwitchD] display isis interface
```

```

                          Interface information for ISIS(1)
                          -----
Interface: Vlan-interface100
Id      IPV4.State      IPV6.State      MTU      Type      DIS
001     Up              Down            1497     L1/L2     No/No

```

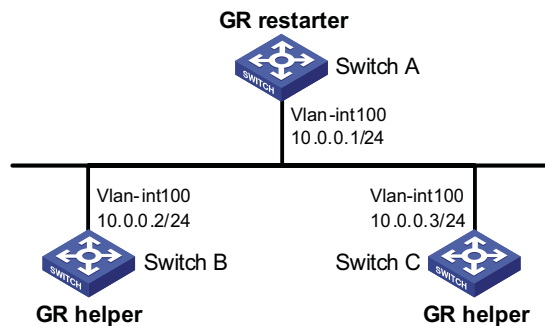
IS-IS Graceful Restart Configuration Example

Network requirements

Switch A, Switch B, and Switch C are interconnected to each other in the same IS-IS routing domain. These switches are reachable to one another through IS-IS, as illustrated in Figure 114.

Network diagram

Figure 114 Network diagram for IS-IS GR configuration



Configuration procedure

1 Configure IP addresses of the interfaces on each switch and configure IS-IS.

Follow Figure 114 to configure the IP address and subnet mask of each interface. The configuration procedure is omitted.

Configure IS-IS on the switches, ensuring that Switch A, Switch B and Switch C can communicate with each other at layer 3 and dynamic route update can be implemented among them with IS-IS. The configuration procedure is omitted here.

2 Configure IS-IS Graceful Restart.

Enable IS-IS Graceful Restart on Switch A and configure the Graceful Restart Interval.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] graceful-restart
[SwitchA-isis-1] graceful-restart interval 150
[SwitchA-isis-1] return
```

Configurations for Switch B and Switch C are similar and therefore are omitted here.

3 Verify the configuration.

After Router A establishes adjacencies with Router B and Router C, they begin to exchange routing information. Restart IS-IS on Router A, which enters into the restart state and sends connection requests to its neighbors through the Graceful Restart mechanism to synchronize the LSDB. Using the **display isis graceful-restart status** command can display the IS-IS GR status on Router A.

Restart Switch A.

```
<SwitchA> reset isis all 1
Warning : Reset ISIS process? [Y/N]:y
```

Check the Graceful Restart status of IS-IS on Switch A.

```
<SwitchA> display isis graceful-restart status
Restart information for ISIS(1)
-----
```

```
IS-IS(1) Level-1 Restart Status
Restart Interval: 150
SA Bit Supported
Total Number of Interfaces = 1
Restart Status: RESTARTING
T3 Timer Status:
  Remaining Time: 65535
T2 Timer Status:
  Remaining Time: 59
Interface Vlan1
  T1 Timer Status:
    Remaining Time: 1
  RA Not Received
```

Complete CSNP Not Received
Number of T1 Pre Expiry: 0

IS-IS(1) Level-2 Restart Status
Restart Interval: 150
SA Bit Supported
Total Number of Interfaces = 1
Restart Status: RESTARTING
T3 Timer Status:
Remaining Time: 65535
T2 Timer Status:
Remaining Time: 59
Interface Vlan1
T1 Timer Status:
Remaining Time: 1
RA Not Received
Complete CSNP Not Received
Number of T1 Pre Expiry: 0

30

BGP CONFIGURATION

The Border Gateway Protocol (BGP) is a dynamic inter-AS route discovery protocol.

When configuring BGP, go to these sections for information you are interested in:

- “BGP Overview” on page 365
- “BGP Configuration Task List” on page 380
- “Configuring BGP Basic Functions” on page 381
- “Controlling Route Distribution and Reception” on page 383
- “Configuring BGP Route Attributes” on page 386
- “Tuning and Optimizing BGP Networks” on page 388
- “Configuring a Large Scale BGP Network” on page 390
- “Configuring BGP GR” on page 392
- “Displaying and Maintaining BGP” on page 394
- “BGP Configuration Examples” on page 395
- “Troubleshooting BGP” on page 413



The term “router” refers to a router in a generic sense or a Layer 3 switch running routing protocols.

BGP Overview

Three early versions of BGP are BGP-1 (RFC1105), BGP-2 (RFC1163) and BGP-3 (RFC1267). The current version in use is BGP-4 (RFC1771). BGP-4 is rapidly becoming the defacto Internet exterior routing protocol standard and is commonly used between ISPs.



BGP refers to BGP-4 in this document.

The characteristics of BGP are as follows:

- Focusing on the control of route propagation and the selection of optimal routes rather than the discovery and calculation of routes, which makes BGP, an exterior routing protocol different from interior routing protocols such as OSPF and RIP
- Using TCP as its transport layer protocol to enhance reliability
- Supporting CIDR
- Substantially reducing bandwidth occupation by advertising updating routes only and applicable to advertising a great amount of routing information on the Internet

- Eliminating route loops completely by adding AS path information to BGP routes
- Providing abundant routing policies to implement flexible route filtering and selection
- Easy to extend, satisfying new network developments

A router advertising BGP messages is called a BGP speaker, which exchanges new routing information with other BGP speakers. When a BGP speaker receives a new route or a route better than the current one from another AS, it will advertise the route to all the other BGP speakers in the local AS.

BGP speakers call each other peers, and several associated peers form a peer group.

BGP runs on a router in one of the following two modes:

- IBGP (Interior BGP)
- EBGP (External BGP)

BGP is called IBGP when it runs within an AS and is called EBGP when it runs between ASs.

Formats of BGP Messages

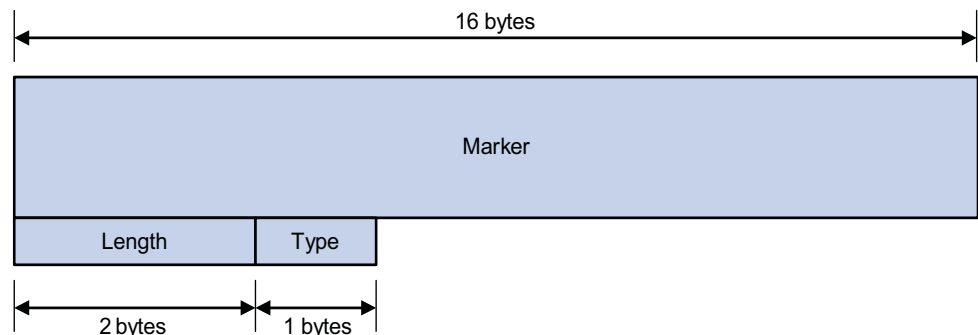
Header

BGP has five types of messages:

- Open
- Update
- Notification
- Keep-alive
- Route-refresh

They have the same header, as shown below:

Figure 115 BGP message header



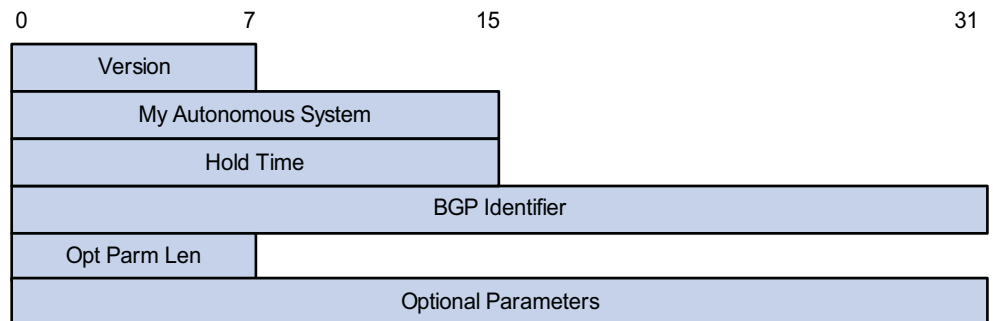
- Marker: The 16-byte field is used for BGP authentication. If no authentication information is available, then the Marker must be all ones.
- Length: The 2-byte unsigned integer indicates the total length of the message.
- Type: This 1-byte unsigned integer indicates the type code of the message. The following type codes are defined: 1-Open, 2-Update, 3-Notification,

4-Keepalive, and 5-Route-refresh. The former four are defined in RFC1771, the last one defined in RFC2918.

Open

After a TCP connection is established, the first message sent by each side is an Open message for peer relationship establishment. The Open message contains the following fields:

Figure 116 BGP open message format

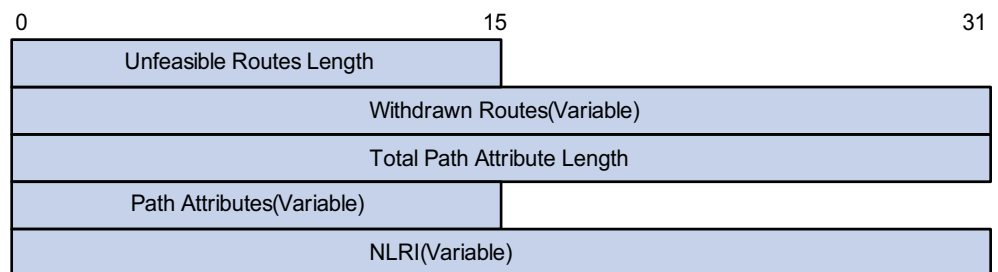


- Version: This 1-byte unsigned integer indicates the protocol version number of the message. The current BGP version is 4.
- My Autonomous System: This 2-byte unsigned integer indicates the Autonomous System number of the sender.
- Hold Time: When establishing peer relationship, two parties negotiate an identical hold time. If no Keepalive or Update is received from a peer after the hold time, the BGP connection is considered down.
- BGP Identifier: In IP address format, identifying the BGP router
- Opt Parm Len (Optional Parameters Length): Length of optional parameters, set to 0 if no optional parameter is available

Update

The Update messages are used to exchange routing information between peers. It can advertise a feasible route or remove multiple unfeasible routes. Its format is shown below:

Figure 117 BGP Update message format



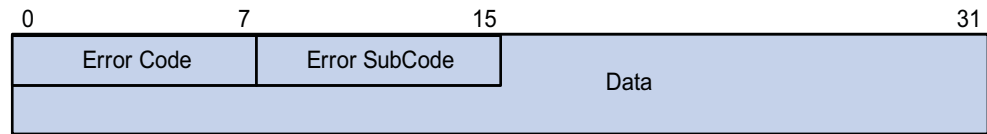
Each Update message can advertise a group of feasible routes with similar attributes, which are contained in the network layer reachable information (NLRI) field. The Path Attributes field carries attributes of these routes that are used by BGP for routing. Each message can also carry multiple withdrawn routes in the Withdrawn Routes field.

- Unfeasible Routes Length: The total length of the Withdrawn Routes field in bytes. A value of 0 indicates neither any route is being withdrawn from service, nor Withdrawn Routes field is present in this Update message.
- Withdrawn Routes: This is a variable length field that contains a list of IP prefixes of routes that are being withdrawn from service.
- Total Path Attribute Length: Total length of the Path Attributes field in bytes. A value of 0 indicates that no Network Layer Reachability Information field is present in this Update message.
- Path Attributes: List of path attributes related to NLRI. Each path attribute is a triple <attribute type, attribute length, attribute value> of variable length. BGP uses these attributes to avoid routing loops, perform routing and protocol extension.
- NLRI (Network Layer Reachability Information): Reachability information is encoded as one or more 2-tuples of the form <length, prefix>.

Notification

A Notification message is sent when an error is detected. The BGP connection is closed immediately after sending it. Notification message format is shown below:

Figure 118 BGP Notification message format



- Error Code: Type of Notification.
- Error Subcode: Specific information about the nature of the reported error.
- Data: Used to diagnose the reason for the Notification. The contents of the Data field depend upon the Error Code and Error Subcode. Erroneous part of data is recorded. The Data field length is variable.

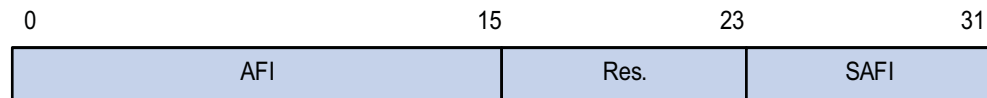
Keepalive

Keepalive messages are sent between peers to maintain connectivity. Its format contains only the message header.

Route-refresh

A route-refresh message is sent to a peer to request the resending of the specified address family routing information. Its format is shown below:

Figure 119 BGP Route-refresh message format



AFI: Address Family Identifier.

Res: Reserved. Set to 0.

SAFI: Subsequent Address Family Identifier.

BGP Path Attributes Classification of path attributes

Path attributes fall into four categories:

- Well-known mandatory: Must be recognized by all BGP routers and must be included in every update message. Routing information error occurs without this attribute.
- Well-known discretionary: Can be recognized by all BGP routers and optional to be included in every update message as needed.
- Optional transitive: Transitive attribute between ASs. A BGP router not supporting this attribute can still receive routes with this attribute and advertise them to other peers.
- Optional non-transitive: If a BGP router does not support this attribute, it will not advertise routes with this attribute.

The usage of each BGP path attribute is described in the following table.

Table 46 Usage of BGP path attributes

Name	Category
ORIGIN	Well-known mandatory
AS_PATH	Well-known mandatory
NEXT_HOP	Well-known mandatory
LOCAL_PREF	Well-known discretionary
ATOMIC_AGGREGATE	Well-known discretionary
AGGREGATOR	Optional transitive
COMMUNITY	Optional transitive
MULTI_EXIT_DISC (MED)	Optional non-transitive
ORIGINATOR_ID	Optional non-transitive
CLUSTER_LIST	Optional non-transitive

Usage of BGP path attributes

1 ORIGIN

ORIGIN is a well-known mandatory attribute and defines the origin of routing information and how a route becomes a BGP route. It involves three types:

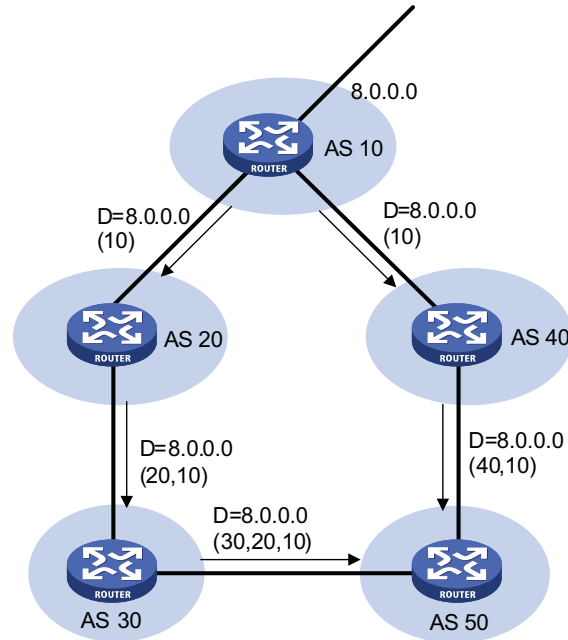
- IGP: Has the highest priority. Routes added to the BGP routing table using the **network** command have the IGP attribute.
- EGP: Has the second highest priority. Routes obtained via EGP have the EGP attribute.
- incomplete: Has the lowest priority. The source of routes with this attribute is unknown, which does not mean such routes are unreachable. The routes redistributed from other routing protocols have the incomplete attribute.

2 AS_PATH

AS_PATH is a well-known mandatory attribute. This attribute identifies the autonomous systems through which routing information carried in this Update message has passed. When a route is advertised from the local AS to another AS, each passed AS number is added into the AS_PATH attribute, thus the receiver can

determine ASs to route the message back. The number of the AS closest to the receiver's AS is leftmost, as shown below:

Figure 120 AS_PATH attribute



In general, a BGP router does not receive routes containing the local AS number to avoid routing loops.



*The current implementation supports using the **peer allow-as-loop** command to receive routes containing the local AS number to meet special requirements.*

The AS_PATH attribute can be used for route selection and filtering. BGP gives priority to the route with the shortest AS_PATH length if other factors are the same. As shown in the above figure, the BGP router in AS50 gives priority to the route passing AS40 for sending information to the destination 8.0.0.0.

In some applications, you can apply a routing policy to control BGP route selection by modifying the AS_PATH length.

By configuring an AS path filtering list, you can filter routes based on AS numbers contained in the AS_PATH attribute.

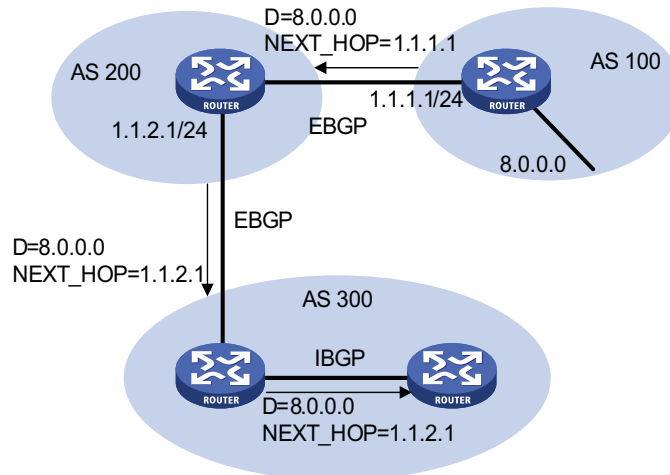
3 NEXT_HOP

Different from IGP, the NEXT_HOP attribute of BGP may not be the IP address of a neighboring router. It involves three types of values, as shown in Figure 121.

- When advertising a self-originated route to an EBGp peer, a BGP speaker sets the NEXT_HOP for the route to the address of its sending interface.
- When sending a received route to an EBGp peer, a BGP speaker sets the NEXT_HOP for the route to the address of the sending interface.
- When sending a route received from an EBGp peer to an IBGP peer, a BGP speaker does not modify the NEXT_HOP attribute. If load-balancing is

configured, the NEXT_HOP attribute will be modified. For load-balancing information, refer to “BGP Route Selection” on page 372.

Figure 121 NEXT_HOP attribute

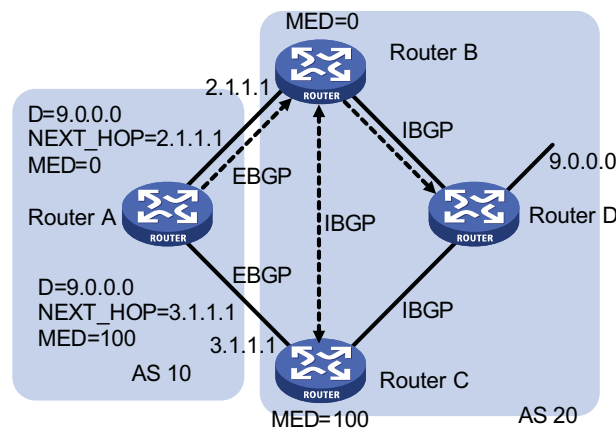


4 MED (MULTI_EXIT_DISC)


The MED attribute is exchanged between two neighboring ASs, each of which does not advertise the attribute to any other AS.

Similar with metrics used by IGP, MED is used to determine the best route for traffic going into an AS. When a BGP router obtains multiple routes to the same destination but with different next hops, it considers the route with the smallest MED value the best route if other conditions are the same. As shown below, traffic from AS10 to AS20 travels through Router B that is selected according to MED.

Figure 122 MED attribute



In general, BGP compares MEDs of routes to the same AS only.

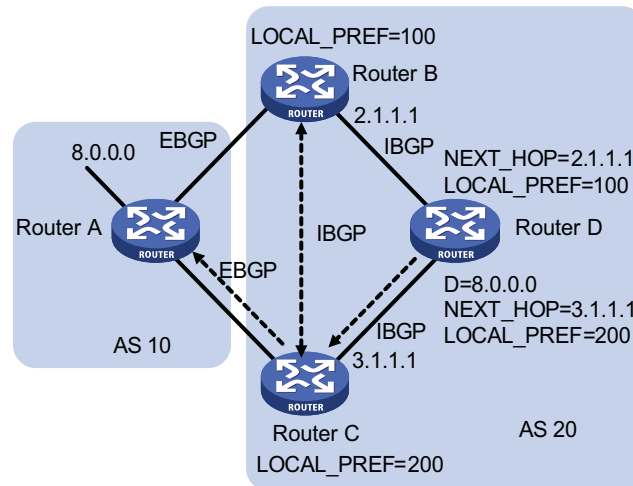
 You can use the **compare-different-as-med** command to force BGP to compare MED values of routes to different ASs.

5 LOCAL_PREF

This attribute is exchanged between IBGP peers only, thus not advertised to any other AS. It indicates the priority of a BGP router.

LOCAL_PREF is used to determine the best route for traffic leaving the local AS. When a BGP router obtains from several IBGP peers multiple routes to the same destination but with different next hops, it considers the route with the highest LOCAL_PREF value as the best route. As shown below, traffic from AS20 to AS10 travels through Router C that is selected according to LOCAL_PREF.

Figure 123 LOCAL_PREF attribute



6 COMMUNITY

The COMMUNITY attribute is used to simplify routing policy usage and ease management and maintenance. It is a collection of destination addresses having identical attributes, without physical boundaries in between, and having nothing to do with the local AS. Well known community attributes include:

- Internet: By default, all routes belong to the Internet community. Routes with this attribute can be advertised to all BGP peers.
- No_Export: After received, routes with this attribute cannot be advertised out the local AS or out the local confederation but can be advertised to other sub-ASs in the confederation (for confederation information, refer to “Settlements for Problems Caused by Large Scale BGP Networks” on page 375).
- No_Advertise: After received, routes with this attribute cannot be advertised to other BGP peers.
- No_Export_Subconfed: After received, routes with this attribute cannot be advertised out the local AS or other ASs in the local confederation.

BGP Route Selection **Route selection rules**

BGP supports the following route selection rules:

- Discard routes with unreachable NEXT_HOP first
- Select the route with the highest Preferred_value
- Select the route with the highest LOCAL_PREF

- Select the route originated by the local router
- Select the route with the shortest AS-PATH
- Select IGP, EGP, Incomplete routes in turn
- Select the route with the lowest MED value
- Select routes learned from EBGP, confederation, IBGP in turn
- Select the route with the smallest next hop cost
- Select the route with the shortest CLUSTER_LIST
- Select the route with the smallest ORIGINATOR_ID
- Select the route advertised by the router with the smallest Router ID



- *CLUSTER_IDs of route reflectors form a CLUSTER_LIST. If a route reflector receives a route that contains its own CLUSTER ID in the CLUSTER_LIST, the router discards the route to avoid routing loops.*
- *If load balancing is configured, the system selects available routes to implement load balancing.*

Route selection with BGP load balancing

The next hop of a BGP route may not be a directly connected neighbor. One of the reasons is next hops in routing information exchanged between IBGPs are not modified. In this case, a router finds the direct route via IGP route entries to reach the next hop. The direct route is called the reliable route. The process of finding a reliable route to reach a next hop is route recursion.

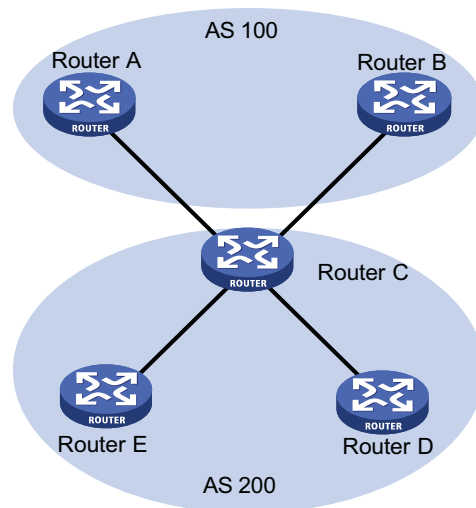
Currently, the switch supports BGP load balancing based on route recursion, namely if reliable routes are load balanced (suppose three next hop addresses), BGP generates the same number of next hops to forward packets. Note that BGP load balancing based on route recursion is always enabled on the switch rather than configured using commands.

BGP differs from IGP in the implementation of load balancing in the following:

- IGP routing protocols such as RIP, OSPF compute metrics of routes, and then implement load balancing on routes with the same metric and to the same destination. The route selection criterion is metric.
- BGP has no route computation algorithm, so it cannot implement load balancing according to metrics of routes. However, BGP has abundant route selection rules, through which, it selects available routes for load balancing and adds load balancing to route selection rules.



- *BGP implements load balancing only on routes that have the same AS_PATH, ORIGIN, LOCAL_PREF and MED.*
- *BGP load balancing is applicable between EBGPs, between IBGPs and between confederations.*
- *If multiple routes to the same destination are available, BGP selects routes for load balancing according to the configured maximum number of load balanced routes.*

Figure 124 Network diagram for BGP load balancing

In the above figure, Router D and Router E are IBGP peers of Router C. Router A and Router B both advertise a route destined for the same destination to Router C. If load balancing is configured and the two routes have the same AS_PATH attribute, ORIGIN attribute, LOCAL_PREF and MED, Router C installs both the two routes to its route table for load balancing. After that, Router C forwards routes to Router D and Router E only once, with AS_PATH unchanged, NEXT_HOP changed to Router C's address. Other BGP transitive attributes apply according to route selection rules.

BGP route advertisement rules

BGP supports the following route advertisement rules:

- When multiple feasible routes exist, a BGP speaker advertises only the best route to its peers.
- A BGP speaker advertises only routes used by itself.
- A BGP speaker advertises routes learned through EBGP to all BGP peers, including both EBGP and IBGP peers.
- A BGP speaker does not advertise IBGP routes to IBGP peers.
- A BGP speaker advertises IBGP routes to EBGP peers. Note that if BGP and IGP synchronization is disabled, IBGP routes are advertised to EBGP peers directly. If the feature is enabled, only IGP advertises the IBGP routes can BGP advertise these routes to EBGP peers.
- A BGP speaker advertises all routes to a newly connected peer.

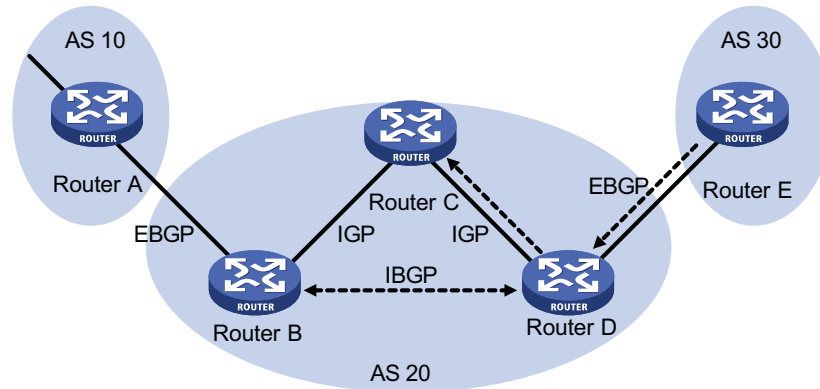
IBGP and IGP Synchronization

The routing information synchronization between IBGP and IGP is for avoidance of giving wrong directions to routers outside of the local AS.

If a non-BGP router works in an AS, a packet forwarded via the router may be discarded due to an unreachable destination. As shown in Figure 125, Router E learned a route of 8.0.0.0/8 from Router D via BGP. Then Router E sends a packet to Router A through Router D, which finds from its routing table that Router B is the next hop (configured using the **peer next-hop-local** command). Since Router D learned the route to Router B via IGP, it forwards the packet to Router C using

route recursion. Router C has no idea about the route 8.0.0.0/8, so it discards the packet.

Figure 125 IBGP and IGP synchronization



If synchronization is configured in this example, the IBGP router (Router D) checks the learned IBGP route from its IGP routing table first. Only the route is available in the IGP routing table can the IBGP router add the route into its BGP routing table and advertise the route to the EBGP peer.

You can disable the synchronization feature in the following cases:

- The local AS is not a transitive AS (AS20 is a transitive AS in the above figure).
- IBGP routers in the local AS are fully meshed.

Settlements for Problems Caused by Large Scale BGP Networks

Route summarization

The size of BGP routing tables on a large network is very large. Using route summarization can reduce the routing table size.

By summarizing multiple routes with one route, a BGP router advertises only the summary route rather than all routes.

Currently, the system supports both manual and automatic summarization. The latter provides for controlling the attribute of a summary route and deciding whether to advertise the route.

Route dampening

BGP route dampening is used to solve the issue of route instability such as route flaps, that is, a route comes up and disappears in the routing table frequently.

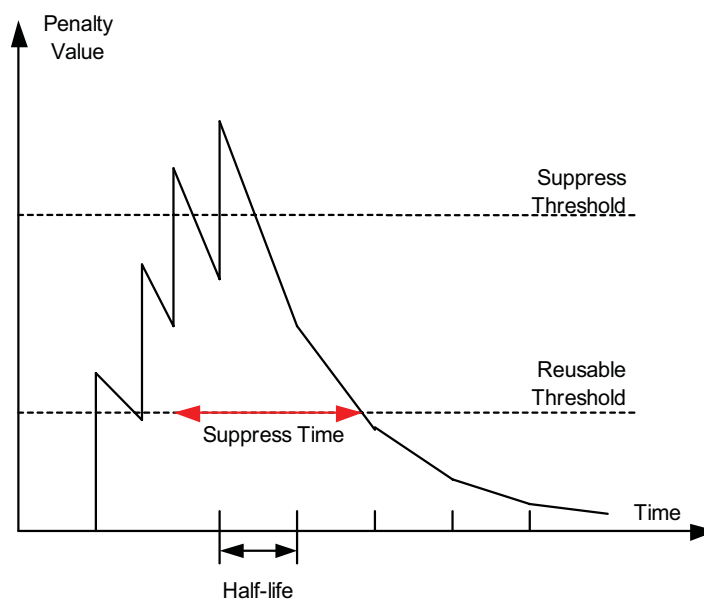
When a route flap occurs, the routing protocol sends an update to its neighbor, and then the neighbor needs to recalculate routes and modify the routing table. Therefore, frequent route flaps consume large bandwidth and CPU resources even affect normal operation of the network.

In most cases, BGP is used in complex networks, where route changes are very frequent. To solve the problem caused by route flaps, BGP uses route dampening to suppress unstable routes.

BGP route dampening uses a penalty value to judge the stability of a route. The bigger the value, the less stable the route. Each time a route flap occurs (the state change of a route from active to inactive is a route flap), BGP adds a penalty value (1000, which is a fixed number and cannot be changed) to the route. When the penalty value of the route exceeds the suppress value, the route is suppressed. That is, it is neither added into the routing table, nor advertised to other BGP peers.

The penalty value of the suppressed route will reduce to half of the suppress value after a period of time. This period is called Half-life. When the value decreases to the reusable threshold value, the route is added into the routing table and advertised to other BGP peers in update packets.

Figure 126 BGP route dampening



Peer group

A peer group is a collection of peers with the same attributes. When a peer joins the peer group, the peer obtains the same configuration as the peer group. If configuration of the peer group is changed, configuration of group members is also changed.

There are many peers in a large BGP network. Some of these peers may be configured with identical commands. The peer group feature simplifies configuration of this kind.

When a peer is added into a peer group, the peer enjoys the same route update policy as the peer group to improve route distribution efficiency.



CAUTION: If an option is configured both for a peer and for the peer group, the latest configuration takes effect.

Community

A peer group makes peers in it enjoy the same policy, while a community makes a group of BGP routers in several ASs enjoy the same policy. Community is a path attribute and advertised between BGP peers, without being limited by AS.

A BGP router can modify the community attribute for a route before sending it to other peers.

Besides using the well-known community attribute, you can define the extended community attribute using a community list to help define a routing policy.

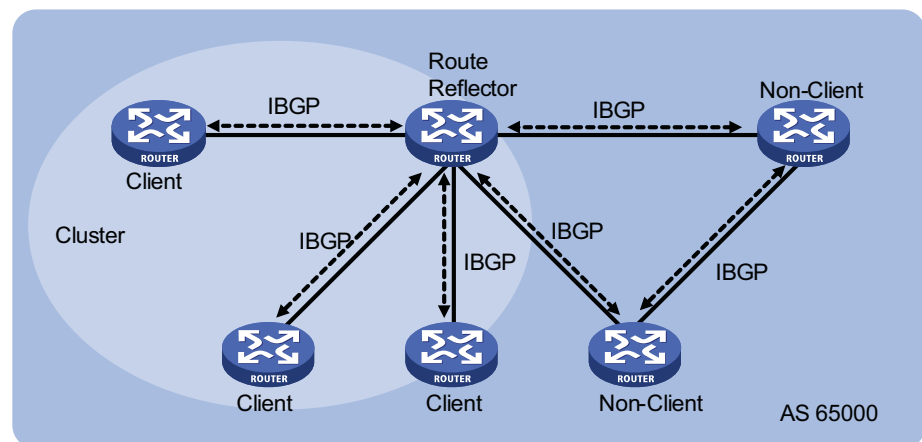
Route reflector

IBGP peers should be fully meshed to maintain connectivity. If there are n routers in an AS, the number of IBGP connections is $n(n-1)/2$. Therefore if there are many IBGP peers, most network and CPU resources will be consumed.

Using route reflectors can solve the issue. In an AS, a router acts as a route reflector, and other routers act as clients connecting to the route reflector. The route reflector forwards (reflects) routing information between clients. BGP connections between clients need not be established.

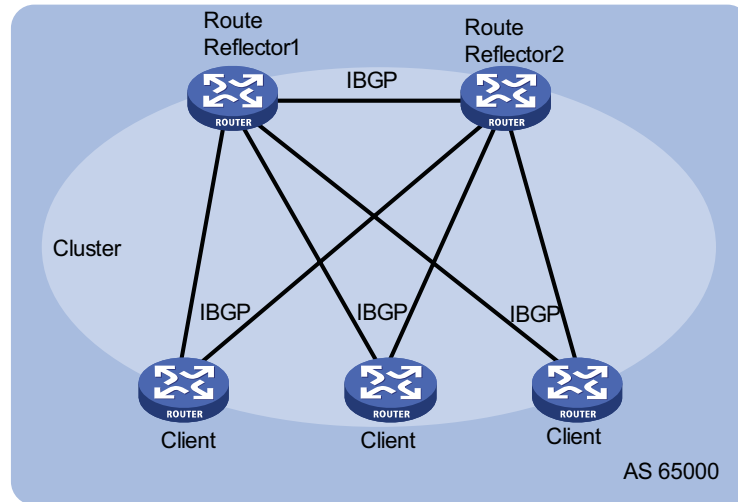
The router neither a route reflector nor a client is a non-client, which has to establish connections to the route reflector and non-clients, as shown below.

Figure 127 Network diagram for route reflector



The route reflector and clients form a cluster. In some cases, you can configure more than one route reflector in a cluster to improve network reliability and prevent the single point failure, as shown in the following figure. The configured route reflectors must have the same Cluster_ID to avoid routing loops.

Figure 128 Network diagram for route reflectors



When clients of a route reflector are fully meshed, route reflection is unnecessary because it consumes more bandwidth resources. The system supports using related commands to disable route reflection in this case.

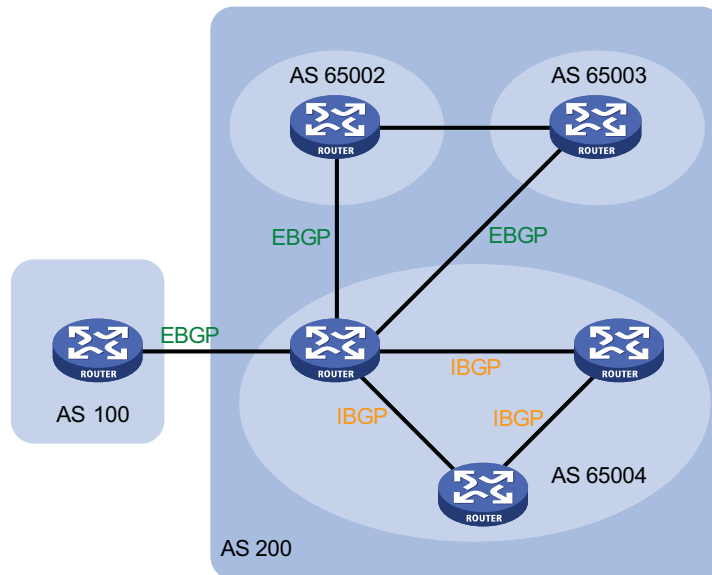


After route reflection is disabled between clients, routes between a client and a non-client can still be reflected.

Confederation

Confederation is another method to deal with growing IBGP connections in ASs. It splits an AS into multiple sub-ASs. In each sub-AS, IBGP peers are fully meshed, and EBGP connections are established between sub-ASs, as shown below:

Figure 129 Confederation network diagram



From the perspective of a non-confederation speaker, it needs not know sub-ASs in the confederation. The ID of the confederation is the number of the AS. In the above figure, AS200 is the confederation ID.

The deficiency of confederation is: when changing an AS into a confederation, you need to reconfigure your routers, and the topology will be changed.

In large-scale BGP networks, both route reflector and confederation can be used.

BGP GR



For GR (Graceful Restart) information, refer to “GR Overview” on page 247.

- 1 To establish a BGP session with a peer, a BGP GR Restarter sends an OPEN message with GR capability to the peer.
- 2 Upon receipt of this message, the peer is aware that the sending router is capable of Graceful Restart, and sends an OPEN message with GR Capability to the GR Restarter to establish a GR session. If neither party has the GR capability, the session established between them will not be GR capable.
- 3 The GR session between the GR Restarter and its peer goes down when the GR Restarter restarts BGP. The GR capable peer will mark all routes associated with the GR Restarter as stale. However, during the configured GR Time, it still uses these routes for packet forwarding, ensuring that no packet will be lost when routing information from its peer is recollected.
- 4 After the restart, the GR Restarter will reestablish a GR session with its peer and send a new GR message notifying the completion of restart. Routing information is exchanged between them for the GR Restarter to create a new routing table and forwarding table with stale routing information removed. Thus the BGP routing convergence is complete.

MP-BGP Overview

The legacy BGP-4 supports IPv4, but does not support other network layer protocols like IPv6.

To support more network layer protocols, IETF extended BGP-4 by introducing Multiprotocol Extensions for BGP-4 (MP-BGP), which is defined in RFC2858.

Routers supporting MP-BGP can communicate with routers not supporting MP-BGP.

MP-BGP extended attributes

In BGP-4, the three types of attributes for IPv4, namely NLRI, NEXT_HOP and AGGREGATOR (contains the IP address of the speaker generating the summary route) are all carried in updates.

To support multiple network layer protocols, BGP-4 puts information about network layer into NLRI and NEXT_HOP. MP-BGP introduced two path attributes:

- MP_REACH_NLRI: Multiprotocol Reachable NLRI, for advertising feasible routes and next hops
- MP_UNREACH_NLRI: Multiprotocol Unreachable NLRI, for withdrawing unfeasible routes

The above two attributes are both optional non-transitive, so BGP speakers not supporting multi-protocol ignore the two attributes and do not forward them to peers.

Address family

MP-BGP employs address family to differentiate network layer protocols. For address family values, refer to RFC 1700 (Assigned Numbers). Currently, the system supports multiple MP-BGP extensions, including IPv6 extension. Different extensions are configured in respective address family view.



- For information about the IPv6 extension application, refer to “IPv6 BGP Configuration” on page 467.
- This chapter gives no detailed commands related to any specific extension application in MP-BGP address family view.

Protocols and Standards

- RFC1771: A Border Gateway Protocol 4 (BGP-4)
- RFC2858: Multiprotocol Extensions for BGP-4
- RFC3392: Capabilities Advertisement with BGP-4
- RFC2918: Route Refresh Capability for BGP-4
- RFC2439: BGP Route Flap Damping
- RFC1997: BGP Communities Attribute
- RFC2796: BGP Route Reflection
- RFC3065: Autonomous System Confederations for BGP
- draft-ietf-idr-restart-08: Graceful Restart Mechanism for BGP

BGP Configuration Task List

Complete the following tasks to configure BGP:

Task	Remarks
“Configuring BGP Basic Functions” on page 381	Required
“Controlling Route Distribution and Reception” on page 383	Optional
“Configuring BGP Route Redistribution” on page 383	Optional
“Configuring BGP Route Summarization” on page 383	Optional
“Advertising a Default Route to a Peer or Peer Group” on page 384	Optional
“Configuring BGP Route Distribution Filtering Policies” on page 384	Optional
“Configuring BGP Route Reception Filtering Policies” on page 385	Optional
“Enabling BGP and IGP Route Synchronization” on page 386	Optional
“Configuring BGP Route Dampening” on page 386	Optional
“Configuring BGP Route Attributes” on page 386	Required
“Tuning and Optimizing BGP Networks” on page 388	Required

Task	Remarks	
"Configuring a Large Scale BGP Network" on page 390	"Configuring BGP Peer Groups" on page 390	Optional
	"Configuring BGP Community" on page 391	Optional
	"Configuring a BGP Route Reflector" on page 392	Optional
	"Configuring a BGP Confederation" on page 392	Optional
"Configuring BGP GR" on page 392	Optional	

Configuring BGP Basic Functions

The section describes BGP basic configuration.



- *This section does not differentiate between BGP and MP-BGP.*
- *Since BGP employs TCP, you need to specify IP addresses of peers, which may not be neighboring routers.*
- *Using logical links can also establish BGP peer relationships.*
- *In general, IP addresses of loopback interfaces are used to improve stability of BGP connections.*

Prerequisites

The neighboring nodes are accessible to each other at the network layer.

Configuration Procedure

Follow these steps to configure BGP basic functions:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable BGP and enter BGP view	bgp <i>as-number</i>	Required Not enabled by default
Specify a Router ID	router-id <i>ip-address</i>	Optional If no IP addresses are configured for loopback interface and other interfaces, the task becomes required.
Specify the AS number of a peer or a peer group	peer { <i>group-name</i> <i>ip-address</i> } as-number <i>as-number</i>	Required Not specified by default
Configure a description for a peer or a peer group	peer { <i>group-name</i> <i>ip-address</i> } description <i>description-text</i>	Optional Not configured by default
Enable IPv4 unicast address family for all peers	default ipv4-unicast	Optional Enabled by default
Enable a peer	peer <i>ip-address</i> enable	Optional Enabled by default
Ignore a peer or peer group	peer { <i>group-name</i> <i>ip-address</i> } ignore	Optional Not ignored by default

To do...		Use the command...	Remarks
Enable the logging of peer state changes	globally	log-peer-change	Optional Enabled by default
	for a peer or peer group	peer { <i>group-name</i> <i>ip-address</i> } log-change	Optional Enabled by default
Specify a preferred value for routes from a peer or peer group		peer { <i>group-name</i> <i>ip-address</i> } preferred-value <i>value</i>	Optional The preferred value defaults to 0.
Specify the source interface for establishing TCP connections to a peer or peer group		peer { <i>group-name</i> <i>ip-address</i> } connect-interface <i>interface-type</i> <i>interface-number</i>	Optional By default, BGP uses the outbound interface of the best route to the BGP peer as the source interface for establishing a TCP connection.
Allow the establishment of EBGP connection to a non directly connected peer/peer group		peer { <i>group-name</i> <i>ip-address</i> } ebgp-max-hop [<i>hop-count</i>]	Optional Not allowed by default. By specifying <i>hop-count</i> , you can specify the max hops for the EBGP connection.



- *It is required to specify for a BGP router a router ID, a 32-bit unsigned integer and the unique identifier of the router in the AS.*
- *You can specify a router ID manually. If not, the system selects an IP address as the router ID. The selection sequence is the highest IP address among loopback interface addresses; if not available, then the highest IP address of interfaces. It is recommended to specify a loopback interface address as the router ID to enhance network reliability. Only when the interface with the selected Router ID or the manual Router ID is deleted will the system select another ID for the router.*
- *You need to create a peer group before configuring it. Refer to “Configuring BGP Peer Groups” on page 390 for creating a peer group.*
- *To establish multiple BGP connections between two routers, you need to specify on the local router the source interfaces for establishing TCP connections to the peers on the peering BGP router respectively; otherwise, the local BGP router may fail to establish TCP connections to the peers when using the outbound interfaces of the best routes as the source interfaces.*
- *In general, direct physical links should be available between EBGP peers. If not, you can use the **peer ebgp-max-hop** command to establish a TCP connection over multiple hops between two peers. You need not use this command for directly connected EBGP peers, which employ loopback interfaces for peer relationship establishment.*
- *If you both reference a routing policy and use the **peer** { *group-name* | *ip-address* } **preferred-value** *value* command to set a preferred value for routes from a peer, the routing policy sets a non-zero preferred value for routes matching it. Other routes not matching the routing policy uses the value set with the command. If the preferred value in the routing policy is zero, the routes matching it will also use the value set with the command. For information about using a routing policy to set a preferred value, refer to the command **peer** { *group-name* | *ip-address* } **route-policy** *route-policy-name* { **export** | **import** } in this document, and the command **apply***

preferred-value preferred-value in Routing Policy Commands of the IP Routing Volume.

Controlling Route Distribution and Reception

Prerequisites Before configuring this task, you have completed BGP basic configuration.

Configuring BGP Route Redistribution BGP can advertise the routing information of the local AS to peering ASs, but it redistributes routing information from IGP into BGP rather than self-finding. During route redistribution, BGP can filter routing information from specific routing protocols.

Follow these steps to configure BGP route redistribution:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	-
Enable BGP to redistribute default route into the BGP routing table	default-route imported	Optional Not enabled by default
Redistribute routes from another routing protocol	import-route <i>protocol</i> [<i>process-id</i> [med <i>med-value</i>] route-policy <i>route-policy-name</i>] *]	Required Not redistributed by default
Inject a network to the BGP routing table	network <i>ip-address</i> [<i>mask</i> <i>mask-length</i>] [short-cut route-policy <i>route-policy-name</i>]	Optional Not injected by default



- The **ORIGIN** attribute of routes redistributed using the **import-route** command is Incomplete.
- The **ORIGIN** attribute of networks advertised into the BGP routing table with the **network** command is IGP. These networks must exist in the local IP routing table, and using a routing policy makes routes control more flexible.

Configuring BGP Route Summarization

To reduce the routing table size on medium and large BGP networks, you need to configure route summarization on peers. BGP supports two summarization modes: automatic and manual.

- Automatic summarization: Summarizes redistributed IGP subnets. With the feature configured, BGP advertises only summary natural networks rather than subnets. The default route and routes injected with the **network** command can not be summarized.
- Manual summarization: Summarizes BGP local routes. The manual summary routes enjoy higher priority than automatic ones.

Follow these steps to configure BGP route summarization:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Enter BGP view		bgp <i>as-number</i>	-
Configure BGP route summarization	Configure automatic route summarization	summary automatic	Required
	Configure manual route summarization	aggregate <i>ip-address</i> { <i>mask</i> <i>mask-length</i> } [as-set attribute-policy <i>route-policy-name</i> detail-suppressed origin-policy <i>route-policy-name</i> suppress-policy <i>route-policy-name</i>]*	No route summarization is configured by default. Choose either as needed; if both are configured, the manual route summarization takes effect.

Advertising a Default Route to a Peer or Peer Group

Follow these steps to advertise a default route to a peer or peer group:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Enter BGP view		bgp <i>as-number</i>	-
Advertise a default route to a peer or peer group		peer { <i>group-name</i> <i>ip-address</i> } default-route-advertise [route-policy <i>route-policy-name</i>]	Required Not advertised by default



*With the **peer default-route-advertise** command executed, the router sends a default route with the next hop being itself to the specified peer/peer group, regardless of whether the default route is available in the routing table.*

Configuring BGP Route Distribution Filtering Policies

Follow these steps to configure BGP route distribution filtering policies:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Enter BGP view		bgp <i>as-number</i>	-

To do...	Use the command...	Remarks
Configure the filtering of outgoing redistributed routes	filter-policy { <i>acl-number</i> ip-prefix <i>ip-prefix-name</i> } export [direct isis <i>process-id</i> ospf <i>process-id</i> rip <i>process-id</i>] static]	Required to choose any; Not configured by default; You can configure a filtering policy as needed;
Reference a routing policy to filter routes to a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } route-policy <i>route-policy-name</i> export	If several filtering policies are configured, they are applied in the following sequence:
Reference an ACL to filter routes to a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } filter-policy <i>acl-number</i> export	<ul style="list-style-type: none"> ■ filter-policy export ■ peer filter-policy export ■ peer as-path-acl export ■ peer ip-prefix export ■ peer route-policy export
Reference an AS path ACL to filter routing information to a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } as-path-acl <i>as-path-acl-number</i> export	
Reference an IP prefix list to filter routing information to a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } ip-prefix <i>ip-prefix-name</i> export	Only routes pass the first policy, can they go to the next, and only routes passing all the configured policies, can they be advertised.

Configuring BGP Route Reception Filtering Policies

Follow these steps to configure BGP route reception filtering policies:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	-
Filter incoming routes with an ACL or IP prefix list	filter-policy { <i>acl-number</i> ip-prefix <i>ip-prefix-name</i> } import	Required to choose any; No route reception filtering is configured by default;
Reference a routing policy to filter routes from a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } route-policy <i>policy-name</i> import	You can configure a filtering policy as needed;
Reference an ACL to filter routing information from a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } filter-policy <i>acl-number</i> import	If several filtering policies are configured, they are applied in the following sequence:
Reference an AS path ACL to filter routing information from a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } as-path-acl <i>as-path-acl-number</i> import	<ul style="list-style-type: none"> ■ filter-policy import ■ peer filter-policy import ■ peer as-path-acl import ■ peer ip-prefix import ■ peer route-policy import
Reference an IP prefix list to filter routing information from a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } ip-prefix <i>ip-prefix-name</i> import	Only routes passing the first policy, can they go to the next, and only routes passing all the configured policies, can they be received.
Specify the maximum number of routes that can be received from a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } route-limit <i>limit</i> [<i>percentage</i>]	The number is unlimited by default.



- Only routes permitted by the specified filtering policies can they be installed into the local BGP routing table.
- Members of a peer group can have different route reception filtering policies from the peer group.

Enabling BGP and IGP Route Synchronization

By default, when a BGP router receives an IBGP route, it only checks the reachability of the route's next hop before advertisement. With BGP and IGP synchronization configured, the BGP router cannot advertise the route to EBGP peers unless the route is also available in the IGP routing table.

Follow these steps to configure BGP and IGP synchronization:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	-
Enable synchronization between BGP and IGP	synchronization	Required Not enabled by default

Configuring BGP Route Dampening

By configuring BGP route dampening, you can suppress unstable routes from neither adding them to the local routing table nor advertising them to BGP peers.

Follow these steps to configure BGP route dampening:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	-
Configure BGP route dampening	dampening [<i>half-life-reachable</i> <i>half-life-unreachable</i> <i>reuse</i> <i>suppress</i> <i>ceiling</i> route-policy <i>route-policy-name</i>] *	Optional Not configured by default

Configuring BGP Route Attributes

Prerequisites Before configuring this task, you have configured BGP basic functions.

Configuration Procedure You can configure BGP route attributes to influence BGP route selection.

Follow these steps to configure BGP route attributes:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	-
Configure preferences for external, internal, local routes	preference { <i>external-preference</i> <i>internal-preference</i> <i>local-preference</i> route-policy <i>route-policy-name</i> }	Optional The default preferences of external, internal, and local routes are 255, 255, and 130 respectively.
Configure the default local preference	default local-preference <i>value</i>	Optional 100 by default

To do...		Use the command...	Remarks
Configure the MED attribute	Configure the default MED value	default med <i>med-value</i>	Optional 0 by default
	Enable the comparison of MED of routes from different ASs	compare-different-as-med	Optional Not enabled by default
	Enable the comparison of MED of routes from each AS	bestroute compare-med	Optional Not enabled by default
	Enable the comparison of MED of routes from confederation peers	bestroute med-confederation	Optional Not enabled by default
Specify the router as the next hop of routes to a peer/peer group		peer { <i>group-name</i> <i>ip-address</i> } next-hop-local	Optional By default, routes to an EBGp peer/peer group take the router as the next hop, while routes to an IBGP peer/peer group do not take the local router as the next hop.
Configure the AS_PATH attribute	Configure repeating times of local AS number in routes from a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } allow-as-loop [<i>number</i>]	Optional The local AS number can not be repeated in routes from the peer/peer group.
	Disable the router from taking AS_PATH as a factor for best route selection	bestroute as-path-neglect	Optional By default, the router takes AS_PATH as a factor for best route selection.
	Specify a fake AS number for a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } fake-as <i>as-number</i>	Optional Not specified by default This command is only applicable to an EBGp peer or peer group.
	Substitute local AS number for the AS number of a peer/peer group in the AS_PATH attribute	peer { <i>group-name</i> <i>ip-address</i> } substitute-as	Optional The substitution is not configured by default.
	Configure BGP to not keep private AS number in AS_PATH of updates to a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } public-as-only	Optional By default, BGP updates carry private AS numbers.



- *Using a routing policy can set preferences for routes matching it. Routes not matching it use the default preferences.*
- *If other conditions are identical, the route with the smallest MED value is selected as the best external route.*

- Using the **peer next-hop-local** command can specify the router as the next hop for routes to a peer/peer group. If BGP load balancing is configured, the router specify itself as the next hop for routes to a peer/peer group regardless of whether the **peer next-hop-local** command is configured.
- In a “third party next hop” network, that is, the two EBGP peers reside in a common broadcast subnet, the BGP router does not specify itself as the next hop for routes to the EBGP peer, unless the **peer next-hop-local** command is configured.
- In general, BGP checks whether the *AS_PATH* attribute of a route from a peer contains the local AS number. If so, it discards the route to avoid routing loops.
- You can specify a fake AS number to hide the real one as needed. The fake AS number applies to routes to EBGP peers only, that is, EBGP peers in other ASs can only find the fake AS number.
- The **peer substitute-as** command is used only in specific networking environments. Inappropriate use of the command may cause routing loops.

Tuning and Optimizing BGP Networks

This task involves the following parts:

1 Configure BGP timers

After establishing a BGP connection, two routers send keepalive messages periodically to each other to keep the connection. If a router receives no keepalive message from the peer after the holdtime elapses, it tears down the connection.

When establishing a BGP connection, the two parties compare their holdtime values, taking the shorter one as the common holdtime.

2 Reset BGP connections

After modifying a route selection policy, you have to reset BGP connections to make the new one take effect, causing short time disconnections. The current BGP implementation supports the route-refresh capability. With this capability enabled on all BGP routers in a network, when a policy is modified on a router, the router advertises a route-refresh message to its peers, which then resend their routing information to the router. Therefore, the local router can perform dynamic route update and apply the new policy without tearing down BGP connections.

If a router not supporting route-refresh exists in the network, you must configure the **peer keep-all-routes** command to save all route updates, and then use the **refresh bgp** command to soft-reset BGP connections, to refresh the BGP routing table and apply the new policy without tearing down BGP connections.

3 Configure BGP authentication

BGP employs TCP as the transport protocol. To enhance security, you can configure BGP to perform MD5 authentication when establishing a TCP connection. BGP MD5 authentication is not for BGP packets. It is used to set passwords for TCP connections. If the authentication fails, the TCP connection can not be established.

Prerequisites Before configuring this task, you have configured BGP basic functions

Configuration Procedure Follow these steps to tune and optimize BGP networks:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Enter BGP view		bgp <i>as-number</i>	-
Configure BGP timers	Configure keepalive interval and holdtime	timer keepalive <i>keepalive hold holdtime</i>	Optional The keepalive interval defaults to 60 seconds, holdtime defaults to 180 seconds.
	Configure keepalive interval and holdtime for a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } timer keepalive <i>keepalive hold holdtime</i>	
Configure the interval for sending the same update to a peer/peer group		peer { <i>group-name</i> <i>ip-address</i> } route-update-interval <i>seconds</i>	Optional The intervals for sending the same update to an IBGP peer and an EBGP peer default to 15 seconds and 30 seconds respectively.
Configure BGP soft reset	Disable BGP route-refresh and multi-protocol extensions for a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } capability-advertise conventional	Optional Enabled by default
	Enable BGP route refresh for a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } capability-advertise route-refresh	Optional Enabled by default
	Keep all original routes from a peer/peer group regardless of whether they pass the inbound filtering policy	peer { <i>group-name</i> <i>ip-address</i> } keep-all-routes	Optional Not kept by default
Return to user view		return	-
Perform manual soft reset on BGP connections		refresh bgp { all <i>ip-address</i> group <i>group-name</i> external internal } { export import }	Required
Enter system view		system-view	-
Enter BGP view		bgp <i>as-number</i>	-
Enable the clearing of the direct EBGP session on any interface that becomes down		ebgp-interface-sensitive	Optional Enabled by default
Enable MD5 authentication when establishing a TCP connection to the peer/peer group		peer { <i>group-name</i> <i>ip-address</i> } password { cipher simple } <i>password</i>	Optional Enabled by default

To do...	Use the command...	Remarks
Configure the number of BGP load balanced routes	balance <i>number</i>	Optional Load balancing is not enabled by default.



- *The maximum keepalive interval should be one third of the holdtime and no less than 1 second. The holdtime is no less than 3 seconds unless it is set to 0.*
- *The intervals set with the **peer timer** command are preferred to those set with the **timer** command.*
- *Use of the **peer keep-all-routes** command saves all routing updates from the peer regardless of whether any filtering policy is configured. The system uses these updates to rebuild the routing table after a soft reset.*
- *Performing BGP soft reset can refresh the routing table and apply the new policy without tearing down BGP sessions.*
- *BGP soft reset requires all routers in the network have the route-refresh capability. If not, you need use the **peer keep-all-routes** command to keep all routing information from a BGP peer to perform soft reset.*

Configuring a Large Scale BGP Network

In a large-scale BGP network, configuration and maintenance become difficult due to large numbers of BGP peers. In this case, configuring peer groups makes management easier and improves route distribution efficiency. Peer group includes IBGP peer group, where peers belong to the same AS, and EBGP peer group, where peers belong to different ASs. If peers in an EBGP group belong to the same external AS, the EBGP peer group is a pure EBGP peer group, and if not, a mixed EBGP peer group.

Configuring BGP community can also help simplify routing policy management, and a community has a much larger management scope than a peer group by controlling routing policies of multiple BGP routers.

To guarantee the connectivity between IBGP peers, you need to make them fully meshed. But it becomes unpractical when there are large numbers of IBGP peers. Configuring route reflectors or confederation can solve it. In a large-scale AS, both of them can be used.

Configuration Prerequisites

Before configuring this task, you have made peering nodes accessible to each other at the network layer.

Configuring BGP Peer Groups

Follow these steps to configure BGP peer groups:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	-

To do...		Use the command...	Remarks
Configure an IBGP peer group	Create an IBGP peer group	group <i>group-name</i> [internal]	Optional
	Add a peer into the IBGP peer group	peer <i>ip-address</i> group <i>group-name</i> [as-number <i>as-number</i>]	You can add multiple peers into the group. The system will create these peers automatically and specify the local AS number as their AS in BGP view.
Configure a pure EBGP peer group	Create an EBGP peer group	group <i>group-name</i> external	Optional
	Specify the AS number for the group	peer <i>group-name</i> as-number <i>as-number</i>	You can add multiple peers into the group. The system will create these peers automatically and specify the local AS number as their AS in BGP view.
	Add a peer into the group	peer <i>ip-address</i> group <i>group-name</i> [as-number <i>as-number</i>]	
Configure a mixed EBGP peer group	Create an EBGP peer group	group <i>group-name</i> external	Optional
	Specify a peer and the AS number for the peer	peer <i>ip-address</i> as-number <i>as-number</i>	You can add multiple peers into the group.
	Add a peer into the group	peer <i>ip-address</i> group <i>group-name</i> [as-number <i>as-number</i>]	



- You need not specify the AS number when creating an IBGP peer group.
- If there are peers in a peer group, you can neither change the AS number of the group nor use the **undo** command to remove the AS number
- You need specify the AS number for each peer in a mixed EBGP peer group respectively.

Configuring BGP Community

Follow these steps to configure BGP community:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Enter BGP view		bgp <i>as-number</i>	-
Advertise the community attribute to a peer/peer group	Advertise the community attribute to a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } advertise-community	Required Not configured by default
	Advertise the extended community attribute to a peer/peer group	peer { <i>group-name</i> <i>ip-address</i> } advertise-ext-community	
Apply a routing policy to routes advertised to a peer/peer group		peer { <i>group-name</i> <i>ip-address</i> } route-policy <i>route-policy-name</i> export	Required Not configured by default



- When configuring BGP community, you need to configure a routing policy to define the community attribute, and apply the routing policy to route advertisement.
- For more information, refer to "Routing Policy Configuration" on page 415.

Configuring a BGP Route Reflector

Follow these steps to configure a BGP route reflector:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	-
Configure the router as a route reflector and specify a peer/peer group as its client	peer { <i>group-name</i> <i>ip-address</i> } reflect-client	Required Not configured by default
Enable route reflection between clients	reflect between-clients	Optional Enabled by default
Configure the cluster ID of the route reflector	reflector cluster-id <i>cluster-id</i>	Optional By default, a route reflector uses its router ID as the cluster ID.



- *In general, it is not required to make clients of a route reflector fully meshed. The route reflector forwards routing information between clients. If clients are fully meshed, you can disable route reflection between clients to reduce routing costs.*
- *In general, a cluster has only one route reflector, and the router ID is used to identify the cluster. You can configure multiple route reflectors to improve network stability. In this case, you need to specify the same cluster ID for these route reflectors to avoid routing loops.*

Configuring a BGP Confederation

Follow these steps to configure a BGP confederation:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	-
Configure a BGP confederation	Configure a confederation ID confederation id <i>as-number</i> Specify sub-ASs contained in the confederation confederation peer-as <i>as-number-list</i>	Required Not configured by default
Enable compatibility with routers not compliant with RFC 3065 in the confederation	confederation nonstandard	Optional Not enabled by default



- *A confederation contains 32 sub-ASs at most. The as-number of a sub-AS takes effect in the confederation only.*
- *If routers not compliant with RFC 3065 exist in the confederation, you can use the **confederation nonstandard** command to make the local router compatible with these routers.*

Configuring BGP GR



A device can act as both a GR Restarter and GR Helper at the same time.

Follow these steps to configure BGP GR:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	-
Enable GR Capability for BGP	graceful-restart	Required Disabled by default
Configure the maximum time allowed for the peer to reestablish a BGP session	graceful-restart timer restart <i>timer</i>	Optional 150 seconds by default
Configure the maximum time to wait for the End-of-RIB marker	graceful-restart timer wait-for-rib <i>timer</i>	Optional 180 seconds by default



- *In general the maximum time allowed for the peer (the GR restarter) to reestablish a BGP session should be less than the Holdtime carried in the OPEN message.*
- *The End-Of-RIB (End of Router-Information-Base) indicates the end of route updates.*

Displaying and Maintaining BGP

Displaying BGP

To do...	Use the command...	Remarks
Display peer group information	display bgp group [<i>group-name</i>]	Available in any view
Display advertised BGP routing information	display bgp network	
Display AS path information	display bgp paths [<i>as-regular-expression</i>]	
Display BGP peer/peer group information	display bgp peer [<i>ip-address</i> { log-info verbose } <i>group-name</i> log-info verbose]	
Display BGP routing information	display bgp routing-table [<i>ip-address</i> [{ <i>mask</i> <i>mask-length</i> } [longer-prefixes]]]	
Display routing information matching the AS path ACL	display bgp routing-table as-path-acl <i>as-path-acl-number</i>	
Display BGP CIDR routing information	display bgp routing-table cidr	
Display BGP routing information matching the specified BGP community	display bgp routing-table community [<i>aa:nn</i> &<1-13>] [no-advertise no-export no-export-subconfed]* [whole-match]	
Display routing information matching a BGP community list	display bgp routing-table community-list { <i>basic-community-list-number</i> [whole-match] <i>adv-community-list-number</i> }&<1-16>	
Display BGP dampened routing information	display bgp routing-table dampened	
Display BGP dampening parameter information	display bgp routing-table dampening parameter	
Display BGP routing information originating from different ASs	display bgp routing-table different-origin-as	
Display BGP routing flap statistics	display bgp routing-table flap-info [regular-expression <i>as-regular-expression</i> as-path-acl <i>as-path-acl-number</i> <i>ip-address</i> [{ <i>mask</i> <i>mask-length</i> } [longer-match]]]	
Display routing information to or from a peer	display bgp routing-table peer <i>ip-address</i> { advertised-routes received-routes } [<i>network-address</i> [<i>mask</i> <i>mask-length</i>]] [statistic]	
Display routing information matching a regular expression	display bgp routing-table regular-expression <i>as-regular-expression</i>	
Display BGP routing statistics	display bgp routing-table statistic	

Resetting BGP Connections

To do...	Use the command...	Remarks
Reset all BGP connections	reset bgp all	Available in user view
Reset the BGP connections to an AS	reset bgp as-number	
Reset the BGP connection to a peer	reset bgp ip-address [flap-info]	
Reset all EBGP connections	reset bgp external	
Reset the BGP connections to a peer group	reset bgp group group-name	
Reset all IBGP connections	reset bgp internal	
Reset all IPv4 unicast BGP connections	reset bgp ipv4 all	

Clearing BGP Information

To do...	Use the command...	Remarks
Clear dampened MBGP routing information and release suppressed routes	reset bgp dampening [ip-address [mask mask-length]]	Available in user view
Clear route flap information	reset bgp flap-info [regexp as-path-regexp as-path-acl as-path-acl-number ip-address [mask mask-length]]	

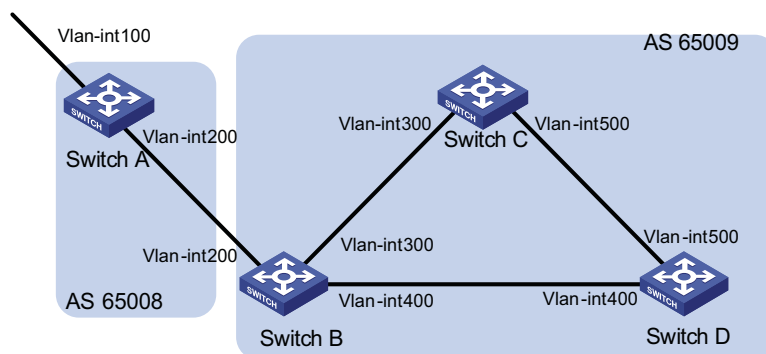
BGP Configuration Examples

BGP Basic Configuration Network requirements

In the following figure are all BGP switches. Between Switch A and Switch B is an EBGP connection. IBGP speakers Switch B, Switch C and Switch D are fully meshed.

Network diagram

Figure 130 Network diagram for BGP basic configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	8.1.1.1/8	Switch D	Vlan-int400	9.1.1.2/24
	Vlan-int200	200.1.1.2/24		Vlan-int500	9.1.2.2/24
Switch B	Vlan-int400	9.1.1.1/24	Switch C	Vlan-int500	9.1.2.1/24
	Vlan-int200	200.1.1.1/24		Vlan-int300	9.1.3.2/24
	Vlan-int300	9.1.3.1/24			

Configuration procedure

- 1 Configure IP addresses for interfaces (omitted)
- 2 Configure IBGP connections

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 65009
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] peer 9.1.1.2 as-number 65009
[SwitchB-bgp] peer 9.1.3.2 as-number 65009
[SwitchB-bgp] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 65009
[SwitchC-bgp] router-id 3.3.3.3
[SwitchC-bgp] peer 9.1.3.1 as-number 65009
[SwitchC-bgp] peer 9.1.2.2 as-number 65009
[SwitchC-bgp] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] bgp 65009
[SwitchD-bgp] router-id 4.4.4.4
[SwitchD-bgp] peer 9.1.1.1 as-number 65009
[SwitchD-bgp] peer 9.1.2.1 as-number 65009
[SwitchD-bgp] quit
```

- 3 Configure the EBGP connection

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 65008
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] peer 200.1.1.1 as-number 65009
```

Inject network 8.0.0.0/8 to the BGP routing table.

```
[SwitchA-bgp] network 8.0.0.0
[SwitchA-bgp] quit
```

Configure Switch B.

```
[SwitchB] bgp 65009
[SwitchB-bgp] peer 200.1.1.2 as-number 65008
[SwitchB-bgp] quit
```

Display BGP peer information on Switch B.

```
[SwitchB] display bgp peer

BGP local router ID : 2.2.2.2
Local AS number : 65009
Total number of peers : 3                Peers in established state : 3
```



```

Peer          V    AS  MsgRcvd  MsgSent  OutQ  PrefRcv  Up/Down  State
9.1.1.2       4 65009    56      56     0      0 00:40:54 Established
9.1.3.2       4 65009    49      62     0      0 00:44:58 Established
200.1.1.2     4 65008    49      65     0      1 00:44:03 Established

```

You can find Switch B has established BGP connections to other switches.

Display BGP routing table information on Switch A.

```

[SwitchA] display bgp routing-table

Total Number of Routes: 1

BGP Local router ID is 1.1.1.1
Status codes: * - valid, > - best, d - damped,
              h - history, i - internal, s - suppressed, S - Stale
              Origin : i - IGP, e - EGP, ? - incomplete
      Network          NextHop          MED          LocPrf    PrefVal Path/Ogn
n
*>  8.0.0.0           0.0.0.0          0              0          i

```

Display BGP routing table information on Switch B.

```

[SwitchB] display bgp routing-table

Total Number of Routes: 1

BGP Local router ID is 2.2.2.2
Status codes: * - valid, > - best, d - damped,
              h - history, i - internal, s - suppressed, S - Stale
              Origin : i - IGP, e - EGP, ? - incomplete
      Network          NextHop          MED          LocPrf    PrefVal Path/Ogn
n
*>  8.0.0.0           200.1.1.2        0              0          65008i

```

Display the BGP routing table on Switch C.

```

[SwitchC] display bgp routing-table

Total Number of Routes: 1

BGP Local router ID is 3.3.3.3
Status codes: * - valid, > - best, d - damped,
              h - history, i - internal, s - suppressed, S - Stale
              Origin : i - IGP, e - EGP, ? - incomplete
      Network          NextHop          MED          LocPrf    PrefVal Path/Ogn
n
i  8.0.0.0           200.1.1.2        0             100        0          65008i

```



From the above outputs, you can find Switch A has learned no route to AS77009, and Switch C has learned network 8.0.0.0 but the next hop 200.1.1.2 is unreachable, so the route is invalid.

4 Redistribute direct routes

Configure Switch B.

```

[SwitchB] bgp 65009
[SwitchB-bgp] import-route direct

```

Display BGP routing table information on Switch A.

```
[SwitchA] display bgp routing-table

Total Number of Routes: 4

BGP Local router ID is 1.1.1.1
Status codes: * - valid, > - best, d - damped,
              h - history, i - internal, s - suppressed, S - Stale
              Origin : i - IGP, e - EGP, ? - incomplete

```

	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>	8.0.0.0	0.0.0.0	0	0		i
*>	9.1.1.0/24	200.1.1.1	0	0		65009?
*>	9.1.3.0/24	200.1.1.1	0	0		65009?
*	200.1.1.0	200.1.1.1	0	0		65009?

Display BGP routing table information on Switch C.

```
[SwitchC] display bgp routing-table

Total Number of Routes: 4

BGP Local router ID is 3.3.3.3
Status codes: * - valid, > - best, d - damped,
              h - history, i - internal, s - suppressed, S - Stale
              Origin : i - IGP, e - EGP, ? - incomplete

```

	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>i	8.0.0.0	200.1.1.2	0	100	0	65008i
*>i	9.1.1.0/24	9.1.3.1	0	100	0	?
* i	9.1.3.0/24	9.1.3.1	0	100	0	?
*>i	200.1.1.0	9.1.3.1	0	100	0	?

You can find the route 8.0.0.0 becomes valid with the next hop being Switch A.

Ping 8.1.1.1 on Switch C.

```
[SwitchC] ping 8.1.1.1
PING 8.1.1.1: 56 data bytes, press CTRL_C to break
  Reply from 8.1.1.1: bytes=56 Sequence=1 ttl=254 time=31 ms
  Reply from 8.1.1.1: bytes=56 Sequence=2 ttl=254 time=47 ms
  Reply from 8.1.1.1: bytes=56 Sequence=3 ttl=254 time=31 ms
  Reply from 8.1.1.1: bytes=56 Sequence=4 ttl=254 time=16 ms
  Reply from 8.1.1.1: bytes=56 Sequence=5 ttl=254 time=31 ms

--- 8.1.1.1 ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 16/31/47 ms
```

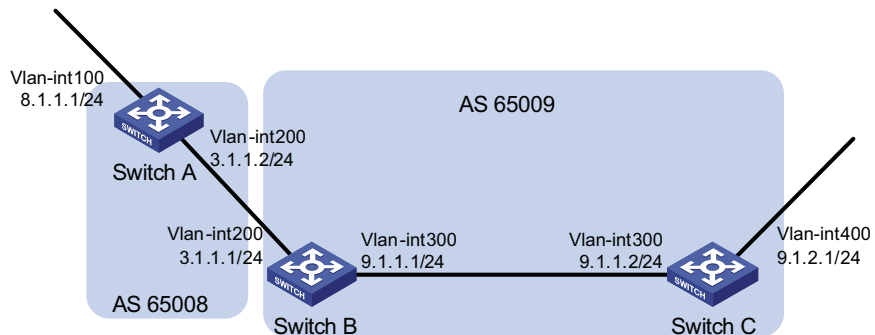
BGP and IGP Synchronization Configuration

Network requirements

As shown below, OSPF is used as the IGP protocol in AS77009, where Switch C is a non-BGP switch. Between Switch A and Switch B is an EBGP connection.

Network diagram

Figure 131 Network diagram for BGP and IGP synchronization



Configuration procedure

- 1 Configure IP addresses for interfaces (omitted)
- 2 Configure OSPF (omitted)
- 3 Configure the EBGP connection

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 65008
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] peer 3.1.1.1 as-number 65009
```

Inject network 8.1.1.0/24 to the BGP routing table.

```
[SwitchA-bgp] network 8.1.1.0 24
[SwitchA-bgp] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 65009
[SwitchB-bgp] peer 3.1.1.2 as-number 65008
[SwitchB-bgp] quit
```

- 4 Configure BGP and IGP synchronization

Configure BGP to redistribute routes from OSPF on Switch B.

```
[SwitchB] bgp 65009
[SwitchB-bgp] import-route ospf 1
[SwitchB-bgp] quit
```

Display routing table information on Switch A.

```
[SwitchA] display bgp routing-table

Total Number of Routes: 3

BGP Local router ID is 1.1.1.1
Status codes: * - valid, > - best, d - damped,
              h - history, i - internal, s - suppressed, S - Stale
Origin : i - IGP, e - EGP, ? - incomplete
```

```

n
      Network          NextHop      MED      LocPrf    PrefVal Path/Og
*>  8.1.1.0/24        0.0.0.0      0          0         i
*>  9.1.1.0/24        3.1.1.1      0          0        65009?
*>  9.1.2.0/24        3.1.1.1      2          0        65009?

```

Configure OSPF to redistribute routes from BGP on Switch B.

```

[SwitchB] ospf
[SwitchB-ospf-1] import-route bgp
[SwitchB-ospf-1] quit

```

Display routing table information on Switch C.

```

<SwitchC> display ip routing-table
Routing Tables: Public
      Destinations : 7          Routes : 7

Destination/Mask    Proto  Pre  Cost      NextHop          Interface
-----
8.1.1.0/24          O_ASE  150  1         9.1.1.1          Vlan300
9.1.1.0/24          Direct  0    0         9.1.1.2          Vlan300
9.1.1.2/32          Direct  0    0         127.0.0.1        InLoop0
9.1.2.0/24          Direct  0    0         9.1.2.1          Vlan400
9.1.2.1/32          Direct  0    0         127.0.0.1        InLoop0
127.0.0.0/8         Direct  0    0         127.0.0.1        InLoop0
127.0.0.1/32        Direct  0    0         127.0.0.1        InLoop0

```

5 Configure route automatic summarization

Configure route automatic summarization on Switch B.

```

[SwitchB] bgp 65009
[SwitchB-bgp] summary automatic

```

Display BGP routing table information on Switch A.

```

[SwitchA] display bgp routing-table

Total Number of Routes: 2

BGP Local router ID is 1.1.1.1
Status codes: * - valid, > - best, d - damped,
              h - history, i - internal, s - suppressed, S - Stale
              Origin : i - IGP, e - EGP, ? - incomplete
      Network          NextHop      MED      LocPrf    PrefVal Path/Ogn
*>  8.1.1.0/24        0.0.0.0      0          0         i
*>  9.0.0.0           3.1.1.1      0          0        65009?

```

Use ping for verification.

```

[SwitchA] ping -a 8.1.1.1 9.1.2.1
PING 9.1.2.1: 56 data bytes, press CTRL_C to break
  Reply from 9.1.2.1: bytes=56 Sequence=1 ttl=254 time=15 ms
  Reply from 9.1.2.1: bytes=56 Sequence=2 ttl=254 time=31 ms
  Reply from 9.1.2.1: bytes=56 Sequence=3 ttl=254 time=47 ms
  Reply from 9.1.2.1: bytes=56 Sequence=4 ttl=254 time=46 ms
  Reply from 9.1.2.1: bytes=56 Sequence=5 ttl=254 time=47 ms

```

```

--- 9.1.2.1 ping statistics ---
 5 packet(s) transmitted
 5 packet(s) received
 0.00% packet loss
 round-trip min/avg/max = 15/37/47 ms

```

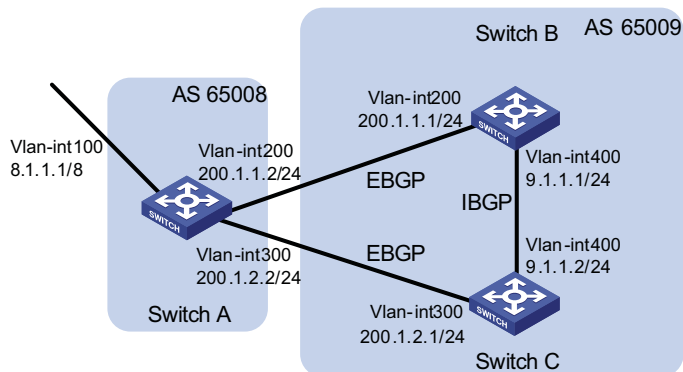
BGP Load Balancing and MED Attribute Configuration

Network requirements

- Configure BGP on all switches; Switch A is in AS77008, and Switch B and C in AS77009.
- Between Switch A and B, and between Switch A and C are EBGP connections, and an IBGP connection is between Switch B and C.

Network diagram

Figure 132 Network diagram for BGP load balancing configuration



Configuration procedure

- 1 Configure IP addresses for interfaces (omitted)
- 2 Configure BGP connections

Configure Switch A.

```

<SwitchA> system-view
[SwitchA] bgp 65008
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] peer 200.1.1.1 as-number 65009
[SwitchA-bgp] peer 200.1.2.1 as-number 65009

```

Inject route 8.0.0.0/8 to BGP routing table.

```

[SwitchA-bgp] network 8.0.0.0 255.0.0.0
[SwitchA-bgp] quit

```

Configure Switch B.

```

<SwitchB> system-view
[SwitchB] bgp 65009
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] peer 200.1.1.2 as-number 65008
[SwitchB-bgp] peer 9.1.1.2 as-number 65009
[SwitchB-bgp] network 9.1.1.0 255.255.255.0
[SwitchB-bgp] quit

```

Configure Switch C.

```

<SwitchC> system-view
[SwitchC] bgp 65009
[SwitchC-bgp] router-id 3.3.3.3
[SwitchC-bgp] peer 200.1.2.2 as-number 65008
[SwitchC-bgp] peer 9.1.1.1 as-number 65009
[SwitchC-bgp] network 9.1.1.0 255.255.255.0
[SwitchC-bgp] quit

```

Display the routing table on Switch A.

```

[SwitchA] display bgp routing-table

Total Number of Routes: 3

BGP Local router ID is 1.1.1.1
Status codes: * - valid, > - best, d - damped,
              h - history, i - internal, s - suppressed, S - Stale
              Origin : i - IGP, e - EGP, ? - incomplete

```

	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>	8.0.0.0	0.0.0.0	0	0	i	
*>	9.1.1.0/24	200.1.1.1	0	0	65009i	
*		200.1.2.1	0	0	65009i	

Two routes to 9.1.1.0/24 are available, and the one with the next hop being 200.1.1.1 is the optimal because the ID of Switch B is smaller.

3 Configure loading balancing**# Configure Switch A.**

```

[SwitchA] bgp 65008
[SwitchA-bgp] balance 2
[SwitchA-bgp] quit

```

Display the routing table on Switch A.

```

[SwitchA] display bgp routing-table

Total Number of Routes: 3

BGP Local router ID is 1.1.1.1
Status codes: * - valid, > - best, d - damped,
              h - history, i - internal, s - suppressed, S - Stale
              Origin : i - IGP, e - EGP, ? - incomplete

```

	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>	8.0.0.0	0.0.0.0	0	0	i	
*>	9.1.1.0/24	200.1.1.1	0	0	65009i	
*>		200.1.2.1	0	0	65009i	

The route 9.1.1.0/24 has two next hops 200.1.1.1 and 200.1.2.1, and both are the optimal.

4 Configure MED**# Configure the default MED of Switch B.**

```

[SwitchB] bgp 65009
[SwitchB-bgp] default med 100

```

Display the routing table on Switch A.

```
[SwitchA] display bgp routing-table

Total Number of Routes: 3

BGP Local router ID is 1.1.1.1
Status codes: * - valid, > - best, d - damped,
              h - history, i - internal, s - suppressed, S - Stale
              Origin : i - IGP, e - EGP, ? - incomplete

```

	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>	8.0.0.0	0.0.0.0	0	0		i
*>	9.1.1.0/24	200.1.2.1	0	0		65009i
*		200.1.1.1	100	0		65009i

From the above information, you can find the route with the next hop 200.1.2.1 is the best route, because its MED (0) is smaller than the MED (100) of the other route with the next hop 200.1.1.1 (Switch B).

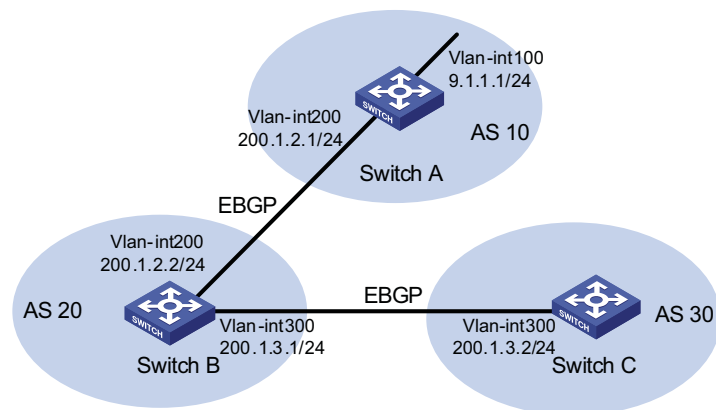
BGP Community Configuration

Network requirements

Switch B establishes EBGP connections with Switch A and C. Configure No_Export community attribute on Switch A to make routes from AS 10 not advertised by AS 20 to any other AS.

Network diagram

Figure 133 Network diagram for BGP community configuration



Configuration procedure

- 1 Configure IP addresses for interfaces (omitted)
- 2 Configure EBGP

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 10
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] peer 200.1.2.2 as-number 20
[SwitchA-bgp] network 9.1.1.0 255.255.255.0
[SwitchA-bgp] quit
```

Configure Switch B.

```

<SwitchB> system-view
[SwitchB] bgp 20
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] peer 200.1.2.1 as-number 10
[SwitchB-bgp] peer 200.1.3.2 as-number 30
[SwitchB-bgp] quit

```

Configure Switch C.

```

<SwitchC> system-view
[SwitchC] bgp 30
[SwitchC-bgp] router-id 3.3.3.3
[SwitchC-bgp] peer 200.1.3.1 as-number 20
[SwitchC-bgp] quit

```

Display the BGP routing table on Switch B.

```
[SwitchB] display bgp routing-table 9.1.1.0
```

```

BGP local router ID : 2.2.2.2
Local AS number : 20
Paths: 1 available, 1 best

```

BGP routing table entry information of 9.1.1.0/24:

```

From          : 200.1.2.1 (1.1.1.1)
Original nexthop: 200.1.2.1
AS-path       : 10
Origin        : igp
Attribute value : MED 0, pref-val 0, pre 255
State         : valid, external, best,
Advertised to such 1 peers:
    200.1.3.2

```

Switch B advertised routes to Switch C in AS30.

Display the routing table on Switch C.

```
[SwitchC] display bgp routing-table
```

```
Total Number of Routes: 1
```

```
BGP Local router ID is 3.3.3.3
```

```
Status codes: * - valid, > - best, d - damped,
```

```
h - history, i - internal, s - suppressed, S - Stale
```

```
Origin : i - IGP, e - EGP, ? - incomplete
```

Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*> 9.1.1.0/24	200.1.3.1		0	20	10i

Switch C learned route 9.1.1.0/24 from Switch B.

3 Configure BGP community

Configure a routing policy.

```

[SwitchA] route-policy comm_policy permit node 0
[SwitchA-route-policy] apply community no-export
[SwitchA-route-policy] quit

```


Apply the routing policy.

```
[SwitchA] bgp 10
[SwitchA-bgp] peer 200.1.2.2 route-policy comm_policy export
[SwitchA-bgp] peer 200.1.2.2 advertise-community
```

Display the routing table on Switch B.

```
[SwitchB] display bgp routing-table 9.1.1.0
BGP local router ID : 2.2.2.2
Local AS number : 20
Paths: 1 available, 1 best

BGP routing table entry information of 9.1.1.0/24:
From          : 200.1.2.1 (1.1.1.1)
Original nexthop: 200.1.2.1
Community     : No-Export
AS-path       : 10
Origin        : igp
Attribute value : MED 0, pref-val 0, pre 255
State         : valid, external, best,
Not advertised to any peers yet
```

The route 9.1.1.0/24 is not available in the routing table of Switch C.

BGP Route Reflector Configuration

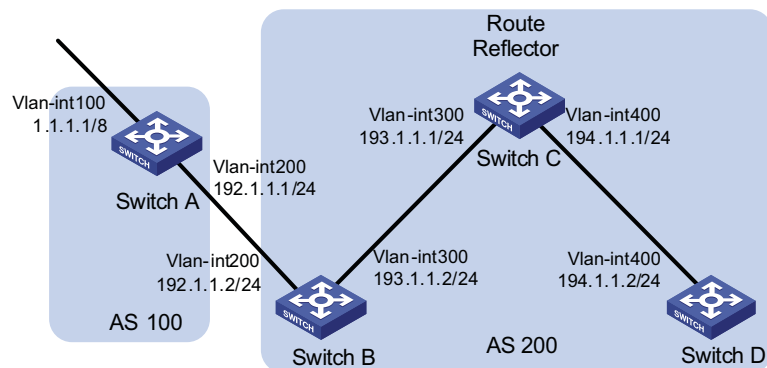
Network requirements

In the following figure, all switches run BGP.

- Between Switch A and Switch B is an EBGP connection, between Switch C and Switch B, and between Switch C and Switch D are IBGP connections.
- Switch C is a route reflector with clients Switch B and D.
- Switch D can learn route 1.0.0.0/8 from Switch C.

Network diagram

Figure 134 Network diagram for BGP route reflector configuration



Configuration procedure

- 1 Configure IP addresses for interfaces (omitted)
- 2 Configure BGP connections

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] peer 192.1.1.2 as-number 200
```

Inject network 1.0.0.0/8 to the BGP routing table.

```
[SwitchA-bgp] network 1.0.0.0
[SwitchA-bgp] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] bgp 200
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] peer 192.1.1.1 as-number 100
[SwitchB-bgp] peer 193.1.1.1 as-number 200
[SwitchB-bgp] peer 193.1.1.1 next-hop-local
[SwitchB-bgp] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] bgp 200
[SwitchC-bgp] router-id 3.3.3.3
[SwitchC-bgp] peer 193.1.1.2 as-number 200
[SwitchC-bgp] peer 194.1.1.2 as-number 200
[SwitchC-bgp] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] bgp 200
[SwitchD-bgp] router-id 4.4.4.4
[SwitchD-bgp] peer 194.1.1.1 as-number 200
[SwitchD-bgp] quit
```

3 Configure the route reflector**# Configure Switch C.**

```
[SwitchC] bgp 200
[SwitchC-bgp] peer 193.1.1.2 reflect-client
[SwitchC-bgp] peer 194.1.1.2 reflect-client
[SwitchC-bgp] quit
```

4 Verify the above configuration**# Display the BGP routing table on Switch B.**

```
[SwitchB] display bgp routing-table
```

```
Total Number of Routes: 1
```

```
BGP Local router ID is 2.2.2.2
```

```
Status codes: * - valid, > - best, d - damped,
```

```
h - history, i - internal, s - suppressed, S - Stale
```

```
Origin : i - IGP, e - EGP, ? - incomplete
```

```

Network      NextHop      MED      LocPrf      PrefVal Path/Ogn
*> 1.0.0.0    192.1.1.1    0        0           100i

```

Display the BGP routing table on Switch D.

```
[SwitchD] display bgp routing-table
```

```
Total Number of Routes: 1
```

```
BGP Local router ID is 4.4.4.4
```

```
Status codes: * - valid, > - best, d - damped,
               h - history, i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
```

```

Network      NextHop      MED      LocPrf      PrefVal Path/Ogn
i 1.0.0.0    193.1.1.2    0        100         0       100i

```

Switch D learned route 1.0.0.0/8 from Switch C.

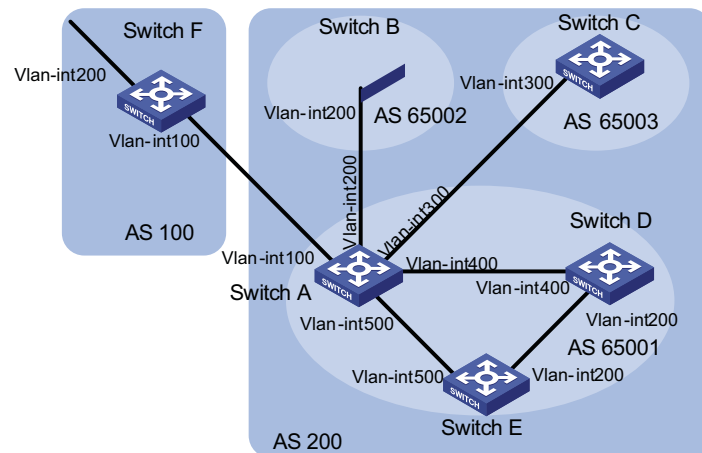
BGP Confederation Configuration

Network requirements

To reduce IBGP connections in AS 200, split it into three sub-ASs, AS77001, AS77002 and AS77003. Switches in AS77001 are fully meshed.

Network diagram

Figure 135 Network diagram for BGP confederation configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	200.1.1.1/24	Switch D	Vlan-int400	10.1.3.2/24
	Vlan-int200	10.1.1.1/24		Vlan-int200	10.1.5.1/24
	Vlan-int300	10.1.2.1/24	Switch E	Vlan-int500	10.1.4.2/24
	Vlan-int400	10.1.3.1/24		Vlan-int200	10.1.5.2/24
	Vlan-int500	10.1.4.1/24	Switch F	Vlan-int200	9.1.1.1/24
Switch B	Vlan-int200	10.1.1.2/24		Vlan-int100	200.1.1.2/24
Switch C	Vlan-int300	10.1.2.2/24			

Configuration procedure

- 1 Configure IP addresses for interfaces (omitted)
- 2 Configure BGP confederation

Configure Switch A.

```

<SwitchA> system-view
[SwitchA] bgp 65001
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] confederation id 200
[SwitchA-bgp] confederation peer-as 65002 65003
[SwitchA-bgp] peer 10.1.1.2 as-number 65002
[SwitchA-bgp] peer 10.1.1.2 next-hop-local
[SwitchA-bgp] peer 10.1.2.2 as-number 65003
[SwitchA-bgp] peer 10.1.2.2 next-hop-local
[SwitchA-bgp] quit

```

Configure Switch B.

```

<SwitchB> system-view
[SwitchB] bgp 65002
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] confederation id 200
[SwitchB-bgp] confederation peer-as 65001 65003
[SwitchB-bgp] peer 10.1.1.1 as-number 65001
[SwitchB-bgp] quit

```

Configure Switch C.

```

<SwitchC> system-view
[SwitchC] bgp 65003
[SwitchC-bgp] router-id 3.3.3.3
[SwitchC-bgp] confederation id 200
[SwitchC-bgp] confederation peer-as 65001 65002
[SwitchC-bgp] peer 10.1.2.1 as-number 65001
[SwitchC-bgp] quit

```

3 Configure IBGP connections in AS77001.

Configure Switch A.

```

[SwitchA] bgp 65001
[SwitchA-bgp] peer 10.1.3.2 as-number 65001
[SwitchA-bgp] peer 10.1.3.2 next-hop-local
[SwitchA-bgp] peer 10.1.4.2 as-number 65001
[SwitchA-bgp] peer 10.1.4.2 next-hop-local
[SwitchA-bgp] quit

```

Configure Switch D.

```

<SwitchD> system-view
[SwitchD] bgp 65001
[SwitchD-bgp] router-id 4.4.4.4
[SwitchD-bgp] confederation id 200
[SwitchD-bgp] peer 10.1.3.1 as-number 65001
[SwitchD-bgp] peer 10.1.5.2 as-number 65001
[SwitchD-bgp] quit

```

Configure Switch E.

```

<SwitchE> system-view
[SwitchE] bgp 65001

```

```
[SwitchE-bgp] router-id 5.5.5.5
[SwitchE-bgp] confederation id 200
[SwitchE-bgp] peer 10.1.4.1 as-number 65001
[SwitchE-bgp] peer 10.1.5.1 as-number 65001
[SwitchE-bgp] quit
```

4 Configure the EBGP connection between AS100 and AS200.

Configure Switch A.

```
[SwitchA] bgp 65001
[SwitchA-bgp] peer 200.1.1.2 as-number 100
[SwitchA-bgp] quit
```

Configure Switch F.

```
<SwitchF> system-view
[SwitchF] bgp 100
[SwitchF-bgp] router-id 6.6.6.6
[SwitchF-bgp] peer 200.1.1.1 as-number 200
[SwitchF-bgp] network 9.1.1.0 255.255.255.0
[SwitchF-bgp] quit
```

5 Verify above configuration

Display the routing table on Switch B.

```
[SwitchB] display bgp routing-table

Total Number of Routes: 1

BGP Local router ID is 2.2.2.2
Status codes: * - valid, > - best, d - damped,
              h - history, i - internal, s - suppressed, S - Stale
              Origin : i - IGP, e - EGP, ? - incomplete
   Network          NextHop      MED       LocPrf   PrefVal Path/Ogn
* > i 9.1.1.0/24    10.1.1.1      0         100      0       (65001) 100i
[SwitchB] display bgp routing-table 9.1.1.0
```

```
BGP local router ID : 2.2.2.2
Local AS number : 65002
Paths: 1 available, 1 best

BGP routing table entry information of 9.1.1.0/24:
From          : 10.1.1.1 (1.1.1.1)
Relay Nexthop : 0.0.0.0
Original nexthop: 10.1.1.1
AS-path       : (65001) 100
Origin        : igp
Attribute value : MED 0, localpref 100, pref-val 0, pre 255
State         : valid, external-confed, best,
Not advertised to any peers yet
```

Display the BGP routing table on Switch D.

```
[SwitchD] display bgp routing-table

Total Number of Routes: 1

BGP Local router ID is 4.4.4.4
Status codes: * - valid, > - best, d - damped,
              h - history, i - internal, s - suppressed, S - Stale
              Origin : i - IGP, e - EGP, ? - incomplete
   Network          NextHop      MED       LocPrf   PrefVal Path/Ogn
```

```
*>i 9.1.1.0/24      10.1.3.1      0      100      0      100i
[SwitchD] display bgp routing-table 9.1.1.0

BGP local router ID : 4.4.4.4
Local AS number : 65001
Paths: 1 available, 1 best

BGP routing table entry information of 9.1.1.0/24:
From      : 10.1.3.1 (1.1.1.1)
Relay Nexthop  : 0.0.0.0
Original nexthop: 10.1.3.1
AS-path     : 100
Origin      : igp
Attribute value : MED 0, localpref 100, pref-val 0, pre 255
State       : valid, internal, best,
Not advertised to any peers yet
```

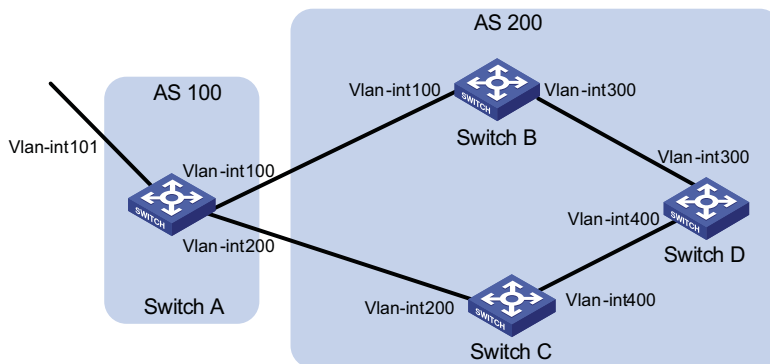
BGP Path Selection Configuration

Network requirements

- In the figure below, all switches run BGP. Between Switch A and Switch B, and between Switch A and Switch C are EBGP connections. Between Switch B and Switch D, and between Switch D and Switch C are IBGP connections.
- OSPF is the IGP protocol in AS 200.
- Configure routing policies, making Switch D use the route 1.0.0.0/8 from Switch C as the optimal.

Network diagram

Figure 136 Network diagram for BGP path selection configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int101	1.0.0.1/8	Switch D	Vlan-int400	195.1.1.1/24
	Vlan-int100	192.1.1.1/24		Vlan-int300	194.1.1.1/24
	Vlan-int200	193.1.1.1/24	Switch C	Vlan-int400	195.1.1.2/24
Switch B	Vlan-int100	192.1.1.2/24		Vlan-int200	193.1.1.2/24
	Vlan-int300	194.1.1.2/24			

Configuration procedure

- 1 Configure IP addresses for interfaces (omitted).
- 2 Configure OSPF on Switch B, C, and D.

Configure Switch B.

```

<SwitchB> system-view
[SwitchB] ospf
[SwitchB-ospf] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 192.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] network 194.1.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] quit

```

Configure Switch C.

```

<SwitchC> system-view
[SwitchC] ospf
[SwitchC-ospf] area 0
[SwitchC-ospf-1-area-0.0.0.0] network 193.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] network 195.1.1.0 0.0.0.255
[SwitchC-ospf-1-area-0.0.0.0] quit
[SwitchC-ospf-1] quit

```

Configure Switch D.

```

<SwitchD> system-view
[SwitchD] ospf
[SwitchD-ospf] area 0
[SwitchD-ospf-1-area-0.0.0.0] network 194.1.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] network 195.1.1.0 0.0.0.255
[SwitchD-ospf-1-area-0.0.0.0] quit
[SwitchD-ospf-1] quit

```

3 Configure BGP connections

Configure Switch A.

```

<SwitchA> system-view
[SwitchA] bgp 100
[SwitchA-bgp] peer 192.1.1.2 as-number 200
[SwitchA-bgp] peer 193.1.1.2 as-number 200

```

Inject network 1.0.0.0/8 to the BGP routing table on Switch A.

```

[SwitchA-bgp] network 1.0.0.0 8
[SwitchA-bgp] quit

```

Configure Switch B.

```

[SwitchB] bgp 200
[SwitchB-bgp] peer 192.1.1.1 as-number 100
[SwitchB-bgp] peer 194.1.1.1 as-number 200
[SwitchB-bgp] quit

```

Configure Switch C.

```

[SwitchC] bgp 200
[SwitchC-bgp] peer 193.1.1.1 as-number 100
[SwitchC-bgp] peer 195.1.1.1 as-number 200
[SwitchC-bgp] quit

```

Configure Switch D.

```
[SwitchD] bgp 200
[SwitchD-bgp] peer 194.1.1.2 as-number 200
[SwitchD-bgp] peer 195.1.1.2 as-number 200
[SwitchD-bgp] quit
```

4 Configure attributes for route 1.0.0.0/8, making Switch D give priority to the route learned from Switch C.

- Configure a higher MED value for the route 1.0.0.0/8 advertised from Switch A to peer 192.1.1.2.

Define an ACL numbered 2000 to permit route 1.0.0.0/8.

```
[SwitchA] acl number 2000
[SwitchA-acl-basic-2000] rule permit source 1.0.0.0 0.255.255.255
[SwitchA-acl-basic-2000] quit
```

Define two routing policies, `apply_med_50`, which sets the MED for route 1.0.0.0/8 to 50, and `apply_med_100`, which sets the MED for route 1.0.0.0/8 to 100.

```
[SwitchA] route-policy apply_med_50 permit node 10
[SwitchA-route-policy] if-match acl 2000
[SwitchA-route-policy] apply cost 50
[SwitchA-route-policy] quit
[SwitchA] route-policy apply_med_100 permit node 10
[SwitchA-route-policy] if-match acl 2000
[SwitchA-route-policy] apply cost 100
[SwitchA-route-policy] quit
```

Apply routing policy `apply_med_50` to the route advertised to peer 193.1.1.2 (Switch C), and `apply_med_100` to the route advertised to peer 192.1.1.2 (Switch B).

```
[SwitchA] bgp 100
[SwitchA-bgp] peer 193.1.1.2 route-policy apply_med_50 export
[SwitchA-bgp] peer 192.1.1.2 route-policy apply_med_100 export
[SwitchA-bgp] quit
```

Display the BGP routing table on Switch D.

```
[SwitchD] display bgp routing-table

Total Number of Routes: 2

BGP Local router ID is 194.1.1.1
Status codes: * - valid, > - best, d - damped,
              h - history, i - internal, s - suppressed, S - Stale
              Origin : i - IGP, e - EGP, ? - incomplete
      Network          NextHop          MED           LocPrf        PrefVal Path/Ogn
-----
*>i 1.0.0.0           193.1.1.1         50             100           0           100i
* i                   192.1.1.1         100            100           0           100i
```

You can find route 1.0.0.0/8 is the optimal.

- Configure different local preferences on Switch B and C for route 1.0.0.0/8, making Switch D give priority to the route from Switch C.

Define an ACL numbered 2000 on Router C, permitting route 1.0.0.0/8.


```
[SwitchC] acl number 2000
[SwitchC-acl-basic-2000] rule permit source 1.0.0.0 0.255.255.255
[SwitchC-acl-basic-2000] quit
```

Configure a routing policy named localpref on Switch C, setting the local preference of route 1.0.0.0/8 to 200 (the default is 100).

```
[SwitchC] route-policy localpref permit node 10
[SwitchC-route-policy] if-match acl 2000
[SwitchC-route-policy] apply local-preference 200
[SwitchC-route-policy] quit
```

Apply routing policy localpref to routes from peer 193.1.1.1.

```
[SwitchC] bgp 200
[SwitchC-bgp] peer 193.1.1.1 route-policy localpref import
[SwitchC-bgp] quit
```

Display the routing table on Switch D.

```
[SwitchD] display bgp routing-table

Total Number of Routes: 2

BGP Local router ID is 194.1.1.1
Status codes: * - valid, > - best, d - damped,
               h - history, i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
      Network          NextHop          MED           LocPrf        PrefVal Path/Ogn
-----
*>i 1.0.0.0           193.1.1.1         0              200           0          100i
* i                   192.1.1.1         0              100           0          100i
```

You can find route 1.0.0.0/8 from Switch D to Switch C is the optimal.

Troubleshooting BGP

No BGP Peer Relationship Established

Symptom

Display BGP peer information using the **display bgp peer** command. The state of the connection to a peer cannot become established.

Analysis

To become BGP peers, any two routers need to establish a TCP session using port 179 and exchange open messages successfully.

Solution

- 1 Use the **display current-configuration** command to verify the peer's AS number.
- 2 Use the **display bgp peer** command to verify the peer's IP address.
- 3 If the loopback interface is used, check whether the **peer connect-interface** command is configured.
- 4 If the peer is a non-direct EBGP peer, check whether the **peer ebgp-max-hop** command is configured.
- 5 Check whether a route to the peer is available in the routing table.

- 6 Use the **ping** command to check connectivity.
- 7 Use the **display tcp status** command to check the TCP connection.
- 8 Check whether an ACL disabling TCP port 179 is configured.



The term “router” refers to a router in a generic sense or a Layer 3 switch running routing protocols.

A routing policy is used on a router for route inspection, filtering, attributes modification when routes are received, advertised, or redistributed.

When configuring routing policy, go to these sections for information you are interested in:

- “Introduction to Routing Policy” on page 415
- “Routing Policy Configuration Task List” on page 417
- “Defining Filtering Lists” on page 417
- “Configuring a Routing Policy” on page 419
- “Displaying and Maintaining the Routing Policy” on page 422
- “Routing Policy Configuration Example” on page 422
- “Troubleshooting Routing Policy Configuration” on page 425



Routing policy described in this chapter includes both IPv4 routing policy and IPv6 routing policy. Configurations of the two are similar, and differences are described in related sections.

Introduction to Routing Policy

Routing Policy and Policy Routing

A routing policy is used on the router for route inspection, filtering, attributes modifying when routes are received, advertised, or redistributed.

Policy routing is a routing mechanism based on the user-defined policies.

This chapter describes only routing policy configuration and usage, refer to “Static Routing Configuration” on page 251 for policy routing information.

When distributing or receiving routing information, a router can use a routing policy to filter routing information. For example, a router receives or advertises only routing information that matches the criteria of a routing policy; a routing protocol redistributes routes from another protocol only routes matching the criteria of a routing policy and modifies some attributes of these routes to satisfy its needs using the routing policy.

To implement a routing policy, you need to define a set of match criteria according to attributes in routing information, such as destination address, advertising

router's address and so on. The match criteria can be set beforehand and then apply them to a routing policy for route distribution, reception and redistribution.

Filters Routing protocols can use six filters: ACL, IP prefix list, AS path ACL, community list, extended community list and routing policy.

ACL

ACL involves IPv4 ACL only. When defining an ACL, you can specify IP addresses and prefixes to match destinations or next hops of routing information.

For ACL configuration, refer to "Configuring an Advanced IPv4 ACL" on page 844.

IP prefix list

IP prefix list plays a role similar to ACL, but it is more flexible than ACL and easier to understand. When an IP prefix list is applied to filtering routing information, its matching object is the destination address of routing information. Moreover, you can specify the **gateway** option to indicate that only routing information advertised by certain routers will be received.

An IP prefix list is identified by name. Each IP prefix list can comprise multiple items, and each item, which is identified by an index number, can specify a matching range in the network prefix format. The index number indicates the matching sequence of items in the IP prefix list.

During matching, the router compares the packet with the items in the ascending order. If one item is matched, the IP prefix list filter is passed, and the packet will not go to the next item.

AS-path

AS path is only applicable to BGP. There is an AS-path field in the BGP packet. An AS path list specifies matching conditions according to the AS-path field.

Community list

Community list only applies to BGP. The BGP packet contains a community attribute field to identify a community. A community list specifies matching conditions based on the community attribute.

Extended community list

Extended community list (extcommunity-list) applies to BGP only. It involves two attributes: Route-Target extcommunity for VPN, Source of Origin extcommunity. An extcommunity-list specifies matching conditions according to the two attributes.

Routing policy

A routing policy is used to match against some attributes in given routing information and modify the attributes of the information if match conditions are satisfied. It can reference the above mentioned filters to define its own match criteria.

A routing policy can comprise multiple nodes, which are in logic OR relationship. Each node is a match unit, and the system compares each node to a packet in the

order of node sequence number. Once a node is matched, the routing policy is passed and the packet will not go through the next node.

Each node comprises a set of **if-match** and **apply** clauses. The **if-match** clauses define the match criteria. The matching objects are some attributes of routing information. The different **if-match** clauses on a node is in logical AND relationship. Only when the matching conditions specified by all the **if-match** clauses on the node are satisfied, can routing information pass the node. The **apply** clauses specify the actions to be performed after the node is passed, concerning the attribute settings for routing information.

Routing Policy Application

A routing policy is applied in two ways:

- When redistributing routes from other routing protocols, a routing protocol accepts only routes passing the routing policy.
- When receiving or advertising routing information, a routing protocol uses the routing policy to filter routing information.

Routing Policy Configuration Task List

Complete the following tasks to configure a routing policy:

Task	
"Defining Filtering Lists" on page 417	"Defining an IPv4 prefix List" on page 417 "Defining an AS Path List" on page 418 "Defining a Community List" on page 418 "Defining an Extended Community List" on page 419
"Configuring a Routing Policy" on page 419	"Creating a Routing Policy" on page 419 "Defining if-match Clauses for the Routing Policy" on page 420 "Defining apply Clauses for the Routing Policy" on page 421

Defining Filtering Lists

Prerequisites

Before configuring this task, you need to decide on:

- IP-prefix list name
- Matching address range
- Extcommunity list sequence number

Defining an IPv4 prefix List

Identified by name, each IPv4 prefix list can comprise multiple items. Each item specifies a matching address range in the form of network prefix identified by index number.

During matching, the system compares the route to each item identified by index number in the ascending order. If one item matches, the route passes the IP-prefix list, without needing to match against the next item.

Follow these steps to define an IPv4 prefix list:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Define an IPv4 prefix list	ip ip-prefix <i>ip-prefix-name</i> [index <i>index-number</i>] { permit deny } <i>ip-address mask-length</i> [greater-equal <i>min-mask-length</i>] [less-equal <i>max-mask-length</i>]	Required Not defined by default



If all items are set to the **deny** mode, no routes can pass the IPv4 prefix list. Therefore, you need to define the **permit 0.0.0.0 0 less-equal 32** item following multiple **deny** mode items to allow other IPv4 routing information to pass.

For example, the following configuration filters routes 10.1.0.0/16, 10.2.0.0/16 and 10.3.0.0/16, but allows other routes to pass.

```
<Sysname> system-view
[Sysname] ip ip-prefix abc index 10 deny 10.1.0.0 16
[Sysname] ip ip-prefix abc index 20 deny 10.2.0.0 16
[Sysname] ip ip-prefix abc index 30 deny 10.3.0.0 16
[Sysname] ip ip-prefix abc index 40 permit 0.0.0.0 0 less-equal 32
```

Defining an AS Path List

You can define multiple items for an AS path ACL that is identified by number. During matching, the relation between items is logical OR, that is, if the route matches one of these items, it passes the AS path ACL.

Follow these steps to define an AS path ACL:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Define an AS path ACL	ip as-path <i>as-path-number</i> { deny permit } <i>regular-expression</i>	Required Not defined by default

Defining a Community List

You can define multiple items for a community list that is identified by number. During matching, the relation between items is logic OR, that is, if routing information matches one of these items, it passes the community list.

Follow these steps to define a community list:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Define a community list	Define a basic community list ip community-list <i>basic-comm-list-num</i> { deny permit } [<i>community-number-list</i>] [internet no-advertise no-export no-export-subconfed] *	Required to define either; Not defined by default
	Define an advanced community list ip community-list <i>adv-comm-list-num</i> { deny permit } <i>regular-expression</i>	

Defining an Extended Community List

You can define multiple items for an extended community list that is identified by number. During matching, the relation between items is logic OR, that is, if routing information matches one of these items, it passes the extended community list.

Follow these steps to define an extended community list:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Define an extended community list	ip extcommunity-list <i>ext-comm-list-number</i> { deny permit } { rt <i>route-target</i> }&<1-16>	Required Not defined by default

Configuring a Routing Policy

A routing policy is used to filter routing information according to some attributes, and modify some attributes of the routing information that matches the routing policy. Match criteria can be configured using filters above mentioned.

A routing policy can comprise multiple nodes, each node contains:

- **if-match** clauses: Define the match criteria that routing information must satisfy. The matching objects are some attributes of routing information.
- **apply** clauses: Specify the actions performed after specified match criteria are satisfied, concerning attribute settings for passed routing information.

Prerequisites

Before configuring this task, you have completed:

- Filtering list configuration
- Routing protocol configuration

You also need to decide on:

- Name of the routing policy, node sequence numbers
- Match criteria
- Attributes to be modified

Creating a Routing Policy

Follow these steps to create a routing policy:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a routing policy and enter its view	route-policy <i>route-policy-name</i> { permit deny } node <i>node-number</i>	Required



- If a node has the **permit** keyword specified, routing information meeting the node's conditions will be handled using the **apply** clauses of this node, without needing to match against the next node. If routing information does not meet the node's conditions, it will go to the next node for a match.
- If a node is specified as **deny**, the **apply** clauses of the node will not be executed. When routing information matches all **if-match** clauses of the node, it can neither pass the node, nor go to the next node. If route information

cannot match any **if-match** clause of the node, it will go to the next node for a match.

- When a routing policy is defined with more than one node, at least one node should be configured with the **permit** keyword. If the routing policy is used to filter routing information, routing information that does not meet any node's conditions cannot pass the routing policy. If all nodes of the routing policy are set using the **deny** keyword, no routing information can pass it.

Defining if-match Clauses for the Routing Policy

Follow these steps to define if-match clauses for a route-policy:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Enter routing policy view		route-policy <i>route-policy-name</i> { permit deny } node <i>node-number</i>	-
Define match criteria for IPv4 routes	Match IPv4 routes having destinations specified in the ACL	if-match acl <i>acl-number</i>	Optional Not configured by default
	Match IPv4 routes having destinations specified in the IP prefix list	if-match ip-prefix <i>ip-prefix-name</i>	
	Match IPv4 routes having next hops or sources specified in the ACL or IP prefix list	if-match ip { next-hop route-source } { acl <i>acl-number</i> ip-prefix <i>ip-prefix-name</i> }	Optional Not configured by default
Match routes having AS path attributes specified in the AS path list (s)		if-match as-path <i>as-path-number</i> &<1-16>	Optional Not configured by default
Match routes having community attributes in the specified community list(s)		if-match community { <i>basic-community-list-number</i> [whole-match] <i>adv-community-list-number</i> }&<1-16>	Optional Not configured by default
Match routes having the specified cost		if-match cost <i>value</i>	Optional Not configured by default
Match BGP routes having extended attributes contained in the extended community list(s)		if-match extcommunity <i>ext-comm-list-number</i> &<1-16>	Optional Not configured by default
Match routes having specified outbound interface(s)		if-match interface { <i>interface-type</i> <i>interface-number</i> }&<1-16>	Optional Not configured by default

To do...	Use the command...	Remarks
Match routes having the specified route type	if-match route-type { internal external-type1 external-type2 external-type1or2 is-is-level-1 is-is-level-2 nssa-external-type1 nssa-external-type2 nssa-external-type1or2 } *	Optional Not configured by default
Match RIP, OSPF, or IS-IS routes having the specified tag value	if-match tag <i>value</i>	Optional Not configured by default



- The **if-match** clauses of a route-policy are in logic AND relationship, namely, routing information has to satisfy all **if-match** clauses before being executed with **apply** clauses.
- You can specify no or multiple **if-match** clauses for a routing policy. If no **if-match** clause is specified, and the routing policy is in permit mode, all routing information can pass the node; if in deny mode, no routing information can pass.

Defining apply Clauses for the Routing Policy

Follow these steps to define apply clauses for a route-policy:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a routing policy and enter its view	route-policy <i>route-policy-name</i> { permit deny } node <i>node-number</i>	Required Not created by default
Set AS_Path attribute for BGP routes	apply as-path <i>as-number</i> &<1-10> [replace]	Optional Not set by default
Specify a community list according to which to delete community attributes of BGP routing information	apply comm-list <i>comm-list-number</i> delete	Optional Not configured by default
Set community attribute for BGP routes	apply community { none additive { <i>community-number</i> &<1-16> <i>aa:nn</i> &<1-16> } internet no-export-subconfed no-export no-advertise } * [additive] }	Optional Not set by default
Set a cost for routes	apply cost [+ -] <i>value</i>	Optional Not set by default
Set a cost type for routes	apply cost-type [external internal type-1 type-2]	Optional Not set by default
Set the extended community attribute for BGP routes	apply extcommunity { rt { <i>as-number:nn</i> <i>ip-address:nn</i> } }&<1-16> [additive] }	Optional Not set by default

To do...	Use the command...	Remarks
Set a next hop for IPv4 routes	apply ip-address next-hop <i>ip-address</i>	Optional Not set by default
Redistribute routes to a specified ISIS level	apply isis { level-1 level-1-2 level-2 }	Optional Not configured by default
Set a local preference for BGP routes	apply local-preference <i>preference</i>	Optional Not set by default
Set an origin attribute for BGP routes	apply origin { igp egp as-number incomplete }	Optional Not set by default
Set a preference for the matched routing protocol	apply preference <i>preference</i>	Optional Not set by default
Set a preferred value for BGP routes	apply preferred-value <i>preferred-value</i>	Optional Not set by default
Set a tag value for RIP, OSPF or IS-IS routes	apply tag <i>value</i>	Optional Not set by default



The **apply ip-address next-hop** command do not apply to redistributed IPv4 routes.

Displaying and Maintaining the Routing Policy

To do...	Use the command...	Remarks
Display BGP AS path ACL information	display ip as-path [<i>as-path-number</i>]	Available in any view
Display BGP community list information	display ip community-list [<i>basic-community-list-number</i> <i>adv-community-list-number</i>]	
Display BGP extended community list information	display ip extcommunity-list [<i>ext-comm-list-number</i>]	
Display IPv4 prefix list statistics	display ip ip-prefix [<i>ip-prefix-name</i>]	
Display routing policy information	display route-policy [<i>route-policy-name</i>]	
Clear IPv4 prefix list statistics	reset ip ip-prefix [<i>ip-prefix-name</i>]	Available in user view

Routing Policy Configuration Example

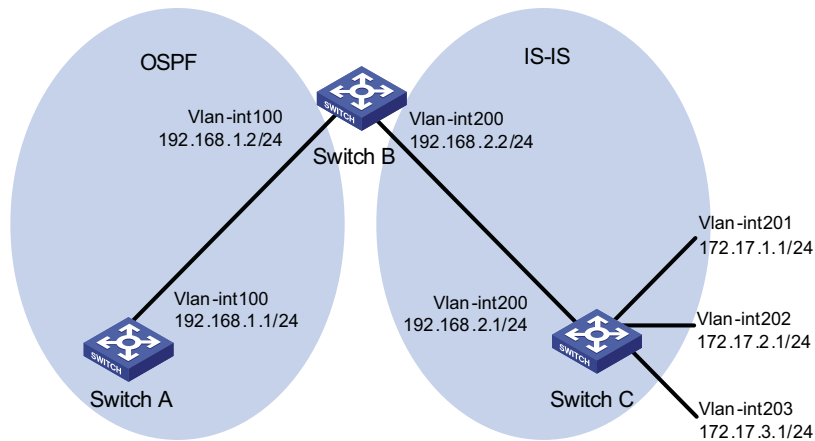
Applying Routing Policy When Redistributing IPv4 Routes

Network Requirements

- Switch B exchanges routing information with Switch A via OSPF, with Switch C via IS-IS.
- On Switch B, configure route redistribution from IS-IS to OSPF and apply a routing policy to set attributes of redistributed routes, setting the cost of route 172.17.1.0/24 to 100, tag of route 172.17.2.0/24 to 20.

Network diagram

Figure 137 Network diagram for routing policy application to route redistribution



Configuration procedure

- 1 Specify IP addresses for interfaces (omitted).
- 2 Configure IS-IS

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis
[SwitchC-isis-1] is-level level-2
[SwitchC-isis-1] network-entity 10.0000.0000.0001.00
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis enable
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 201
[SwitchC-Vlan-interface201] isis enable
[SwitchC-Vlan-interface201] quit
[SwitchC] interface vlan-interface 202
[SwitchC-Vlan-interface202] isis enable
[SwitchC-Vlan-interface202] quit
[SwitchC] interface vlan-interface 203
[SwitchC-Vlan-interface203] isis enable
[SwitchC-Vlan-interface203] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis
[SwitchB-isis-1] is-level level-2
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis enable
[SwitchB-Vlan-interface200] quit
```

- 3 Configure OSPF and route redistribution

Configure Switch A: enable OSPF.

```

<SwitchA> system-view
[SwitchA] ospf
[SwitchA-ospf-1] area 0
[SwitchA-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
[SwitchA-ospf-1-area-0.0.0.0] quit
[SwitchA-ospf-1] quit

```

Configure Switch B: enable OSPF and redistribute routes from IS-IS.

```

[SwitchB] ospf
[SwitchB-ospf-1] area 0
[SwitchB-ospf-1-area-0.0.0.0] network 192.168.1.0 0.0.0.255
[SwitchB-ospf-1-area-0.0.0.0] quit
[SwitchB-ospf-1] import-route isis 1
[SwitchB-ospf-1] quit

```

Display OSPF routing table on Switch A to view redistributed routes.

```

[SwitchA] display ospf routing

                OSPF Process 1 with Router ID 192.168.1.1
                Routing Tables

Routing for Network
Destination      Cost      Type      NextHop      AdvRouter      Area
192.168.1.0/24  1562     Stub     192.168.1.1  192.168.1.1    0.0.0.0

Routing for ASEs
Destination      Cost      Type      Tag      NextHop      AdvRouter
172.17.1.0/24    1         Type2    1         192.168.1.2  192.168.2.2
172.17.2.0/24    1         Type2    1         192.168.1.2  192.168.2.2
172.17.3.0/24    1         Type2    1         192.168.1.2  192.168.2.2
192.168.2.0/24   1         Type2    1         192.168.1.2  192.168.2.2

Total Nets: 5
Intra Area: 1  Inter Area: 0  ASE: 4  NSSA: 0

```

4 Configure filtering lists

Configure an ACL with the number of 2002, letting pass route 172.17.2.0/24.

```

[SwitchB] acl number 2002
[SwitchB-acl-basic-2002] rule permit source 172.17.2.0 0.0.0.255
[SwitchB-acl-basic-2002] quit

```

Configure an IP prefix list named prefix-a, letting pass route 172.17.1.0/24.

```

[SwitchB] ip ip-prefix prefix-a index 10 permit 172.17.1.0 24

```

5 Configure a routing policy.

```

[SwitchB] route-policy isis2ospf permit node 10
[SwitchB-route-policy] if-match ip-prefix prefix-a
[SwitchB-route-policy] apply cost 100
[SwitchB-route-policy] quit
[SwitchB] route-policy isis2ospf permit node 20
[SwitchB-route-policy] if-match acl 2002
[SwitchB-route-policy] apply tag 20
[SwitchB-route-policy] quit
[SwitchB] route-policy isis2ospf permit node 30
[SwitchB-route-policy] quit

```

6 Apply the routing policy to route redistribution.

Configure Switch B: apply the routing policy when redistributing routes.

```
[SwitchB] ospf
[SwitchB-ospf-1] import-route isis 1 route-policy isis2ospf
[SwitchB-ospf-1] quit
```

Display the OSPF routing table on Switch A. You can find the cost of route 172.17.1.0/24 is 100, tag of route 172.17.1.0/24 is 20, and other external routes have no change.

```
[SwitchA] display ospf routing

          OSPF Process 1 with Router ID 192.168.1.1
          Routing Tables

Routing for Network
Destination      Cost      Type      NextHop      AdvRouter      Area
192.168.1.0/24   1         Transit   192.168.1.1  192.168.1.1    0.0.0.0

Routing for ASEs
Destination      Cost      Type      Tag           NextHop      AdvRouter
172.17.1.0/24   100      Type2     1             192.168.1.2  192.168.2.2
172.17.2.0/24   1        Type2     20            192.168.1.2  192.168.2.2
172.17.3.0/24   1        Type2     1             192.168.1.2  192.168.2.2
192.168.2.0/24   1        Type2     1             192.168.1.2  192.168.2.2

Total Nets: 5
Intra Area: 1  Inter Area: 0  ASE: 4  NSSA: 0
```

Troubleshooting Routing Policy Configuration

IPv4 Routing Information Filtering Failure

Symptom

Filtering routing information failed, while routing protocol runs normally.

Analysis

At least one item of the IP prefix list should be configured as permit mode, and at least one node in the Route-policy should be configured as permit mode.

Processing procedure

- 1 Use the **display ip ip-prefix** command to display IP prefix list information.
- 2 Use the **display route-policy** command to display routing policy information.

32

IPv6 STATIC ROUTING CONFIGURATION



The term “router” in this document refers to a Layer 3 switch running routing protocols.

Introduction to IPv6 Static Routing

Static routes are special routes that are manually configured by network administrators. They work well in simple networks. Configuring and using them properly can improve the performance of networks and guarantee enough bandwidth for important applications.

However, static routes also have shortcomings: any topology changes could result in unavailable routes, requiring the network administrator to manually configure and modify the static routes.

Features of IPv6 Static Routes

Similar to IPv4 static routes, IPv6 static routes work well in simple IPv6 network environments.

Their major difference lies in the destination and next hop addresses. IPv6 static routes use IPv6 addresses whereas IPv4 static routes use IPv4 addresses.

Default IPv6 Route

The IPv6 static route that has the destination address configured as `::/0` (indicating a prefix length of 0) is the default IPv6 route. If the destination address of an IPv6 packet does not match any entry in the routing table, this default route will be used to forward the packet.

Configuring an IPv6 Static Route

In small IPv6 networks, IPv6 static routes can be used to forward packets. In comparison to dynamic routes, it helps to save network bandwidth.

Configuration prerequisites

- Enabling IPv6 packet forwarding
- Ensuring that the neighboring nodes are IPv6 reachable

Configuring an IPv6 Static Route

Follow these steps to configure an IPv6 static route:

To do...	Use the commands...	Remarks
Enter system view	System-view	-
Configure an IPv6 static route	ipv6 route-static <i>ipv6-address prefix-length</i> [<i>interface-type interface-number</i>] <i>nexthop-address</i> [preference <i>preference-value</i>]	Required The default preference of IPv6 static routes is 60.

Displaying and Maintaining IPv6 Static Routes

To do...	Use the command...	Remarks
Display IPv6 static route information	display ipv6 routing-table protocol static [inactive verbose]	Available in any view
Remove all IPv6 static routes	delete ipv6 static-routes all	Available in system view



Using the **undo ipv6 route-static** command can delete a single IPv6 static route, while using the **delete ipv6 static-routes all** command deletes all IPv6 static routes including the default route.

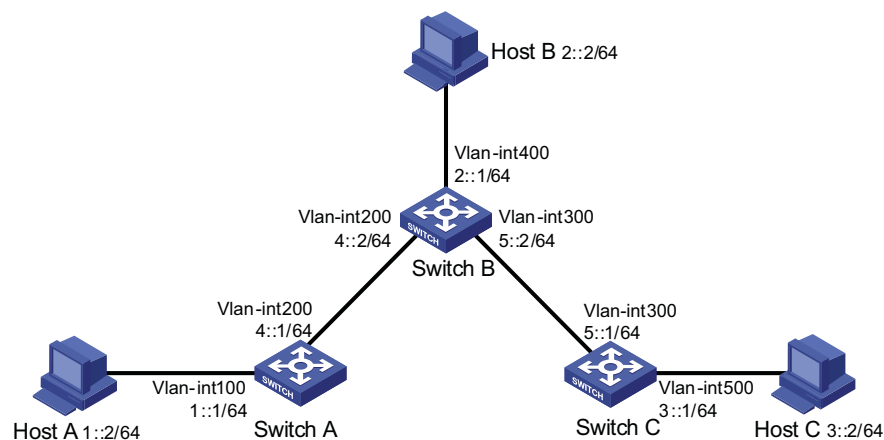
IPv6 Static Routing Configuration Example

Network requirements

With IPv6 static routes configured, all hosts and switches can interact with each other.

Network diagram

Figure 138 Network diagram for static routes



Configuration procedure

- 1 Configure the IPv6 addresses of all VLAN interfaces (Omitted)
- 2 Configure IPv6 static routes.

Configure the default IPv6 static route on Switch A.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] ipv6 route-static :: 0 4::2
```

Configure two IPv6 static routes on Switch B.

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] ipv6 route-static 1:: 64 4::1
[SwitchB] ipv6 route-static 3:: 64 5::1
```


Configure the default IPv6 static route on Switch C.

```
<SwitchC> system-view
[SwitchC] ipv6
[SwitchC] ipv6 route-static :: 0 5::2
```

3 Configure the IPv6 addresses of hosts and gateways.

Configure the IPv6 addresses of all the hosts based upon the network diagram, configure the default gateway of Host A as 1::1, that of Host B as 2::1, and that of Host C as 3::1.

4 Display configuration information

Display the IPv6 routing table of Switch A.

```
[SwitchA] display ipv6 routing-table
Routing Table :
      Destinations : 7          Routes : 7

Destination: ::/0
NextHop      : 4::2
Interface    : Vlan200
Protocol     : Static
Preference   : 60
Cost        : 0

Destination: ::1/128
NextHop      : ::1
Interface    : InLoop0
Protocol     : Direct
Preference   : 0
Cost        : 0

Destination: 1::/64
NextHop      : 1::1
Interface    : Vlan100
Protocol     : Direct
Preference   : 0
Cost        : 0

Destination: 1::1/128
NextHop      : ::1
Interface    : InLoop0
Protocol     : Direct
Preference   : 0
Cost        : 0

Destination: 4::/64
NextHop      : 4::1
Interface    : Vlan200
Protocol     : Direct
Preference   : 0
Cost        : 0

Destination: 4::1/128
NextHop      : ::1
Interface    : InLoop0
Protocol     : Direct
Preference   : 0
Cost        : 0

Destination: FE80::/10
NextHop      : ::
Interface    : NULL0
Protocol     : Direct
Preference   : 0
Cost        : 0
```

Verify the connectivity with the **ping** command.

```
[SwitchA] ping ipv6 3::1
PING 3::1 : 56 data bytes, press CTRL_C to break
  Reply from 3::1
    bytes=56 Sequence=1 hop limit=254  time = 63 ms
  Reply from 3::1
    bytes=56 Sequence=2 hop limit=254  time = 62 ms
  Reply from 3::1
    bytes=56 Sequence=3 hop limit=254  time = 62 ms
  Reply from 3::1
    bytes=56 Sequence=4 hop limit=254  time = 63 ms
  Reply from 3::1
    bytes=56 Sequence=5 hop limit=254  time = 63 ms
```

```
--- 3::1 ping statistics ---  
 5 packet(s) transmitted  
 5 packet(s) received  
 0.00% packet loss  
 round-trip min/avg/max = 62/62/63 ms
```

33

IPv6 RIPNG CONFIGURATION



- The term “router” in this document refers to a Layer 3 switch running routing protocols.
- The Switch 4800G only support single RIPng process.

Introduction to RIPng

RIP next generation (RIPng) is an extension of RIP-2 for IPv4. Most RIP concepts are applicable in RIPng.

RIPng for IPv6 made the following changes to RIP:

- UDP port number: RIPng uses UDP port 521 for sending and receiving routing information.
- Multicast address: RIPng uses FF02:9 as the link-local multicast address.
- Destination Prefix: 128-bit destination address prefix.
- Next hop: 128-bit IPv6 address.
- Source address: RIPng uses FE80::/10 as the link-local source address

RIPng Working Mechanism

RIPng is a routing protocol based on the distance vector (D-V) algorithm. RIPng uses UDP packets to exchange routing information through port 521.

RIPng uses a hop count to measure the distance to a destination. The hop count is referred to as metric or cost. The hop count from a router to a directly connected network is 0. The hop count between two directly connected routers is 1. When the hop count is greater than or equal to 16, the destination network or host is unreachable.

By default, the routing update is sent every 30 seconds. If the router receives no routing updates from a neighbor after 180 seconds, the routes learned from the neighbor are considered as unreachable. After another 240 seconds, if no routing update is received, the router will remove these routes from the routing table.

RIPng supports Split Horizon and Poison Reverse to prevent routing loops, and route redistribution.

Each RIPng router maintains a routing database, including route entries of all reachable destinations. A route entry contains the following information:

- Destination address: IPv6 address of a host or a network.
- Next hop address: IPv6 address of a neighbor along the path to the destination.
- Egress interface: Outbound interface that forwards IPv6 packets.
- Metric: Cost from the local router to the destination.

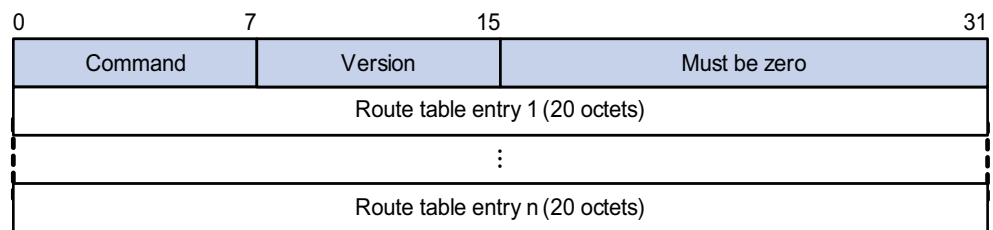
- Route time: Time that elapsed since a route entry is last changed. Each time a route entry is modified, the routing time is set to 0.
- Route tag: Identifies the route, used in routing policy to control routing information.

RIPng Packet Format Basic format

A RIPng packet consists of a header and multiple route table entries (RTEs). The maximum number of RTEs in a packet depends on the MTU of the sending interface.

Figure 139 shows the packet format of RIPng.

Figure 139 RIPng basic packet format



- Command: Type of message. 0x01 indicates Request, 0x02 indicates Response.
- Version: Version of RIPng. It can only be 0x01 currently.
- RTE: Route table entry, 20 bytes for each entry.

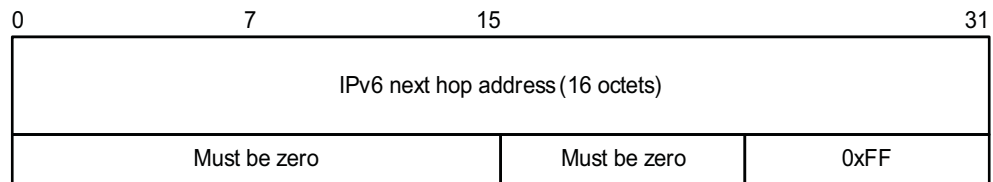
RTE format

There are two types of RTE in RIPng.

- Next hop RTE: Defines the IPv6 address of a next hop
- IPv6 prefix RTE: Describes the destination IPv6 address, route tag, prefix length and metric in the RIPng routing table.

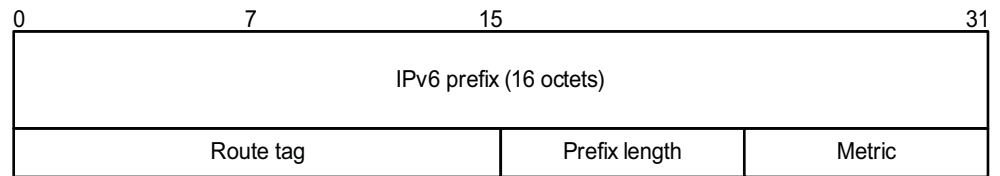
Figure 140 shows the format of the next hop RTE:

Figure 140 Next hop RTE format



IPv6 next hop address is the IPv6 address of the next hop.

Figure 141 shows the format of the IPv6 prefix RTE.

Figure 141 IPv6 prefix RTE format

- IPv6 prefix: Destination IPv6 address prefix.
- Route tag: Route tag.
- Prefix len: Length of the IPv6 address prefix.
- Metric: Cost of a route.

RIPng Packet Processing Procedure

Request packet

When a RIPng router first starts or needs to update some entries in its routing table, generally a multicast request packet is sent to ask for needed routes from neighbors.

The receiving RIPng router processes RTEs in the request. If there is only one RTE with the IPv6 prefix and prefix length both being 0, and with a metric value of 16, the RIPng router will respond with the entire routing table information in response messages. If there are multiple RTEs in the request message, the RIPng router will examine each RTE, update its metric, and send the requested routing information to the requesting router in the response packet.

Response packet

The response packet containing the local routing table information is generated as:

- A response to a request
- An update periodically
- A triggered update caused by route change

After receiving a response, a router checks the validity of the response before adding the route to its routing table, such as whether the source IPv6 address is the link-local address, whether the port number is correct. The response packet failed the check will be discarded.

Protocols and Standards

- RFC2080: RIPng for IPv6
- RFC2081: RIPng Protocol Applicability Statement
- RFC2453: RIP Version 2

Configuring RIPng Basic Functions

In this section, you are presented with the information to configure the basic RIPng features.

You need to enable RIPng first before configuring other tasks, but it is not necessary for RIPng related interface configurations, such as assigning an IPv6 address.

Configuration Prerequisites

Before the configuration, accomplish the following tasks first:

- Enable IPv6 packet forwarding.
- Configure an IP address for each interface, and make sure all nodes are reachable.

Configuration Procedure

Follow these steps to configure the basic RIPng functions:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Create a RIPng process and enter RIPng view	ripng [<i>process-id</i>]	Required Not created by default
Return to system view	quit	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--
Enable RIPng on the interface	ripng process-id enable	Required Disabled by default



If RIPng is not enabled on an interface, the interface will not send and receive any RIPng route.

Configuring RIPng Route Control

Before the configuration, accomplish the following tasks first:

- Configure an IPv6 address on each interface, and make sure all nodes are reachable.
- Configure RIPng basic functions
- Define an IPv6 ACL before using it for route filtering. Refer to “IPv6 ACL Configuration” on page 851 for related information.
- Define an IPv6 address prefix list before using it for route filtering. Refer to section “Defining an IPv6 Prefix List” on page 491 for related information.

Configuring an Additional Routing Metric

An additional routing metric can be added to the metric of an inbound or outbound RIP route, namely, the inbound and outbound additional metric.

The outbound additional metric is added to the metric of a sent route, the route’s metric in the routing table is not changed.

The inbound additional metric is added to the metric of a received route before the route is added into the routing table, so the route’s metric is changed.

Follow these steps to configure an inbound/outbound additional routing metric:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--
Specify an inbound routing additional metric	ripng metricin <i>value</i>	Optional 0 by default

To do...	Use the command...	Remarks
Specify an outbound routing additional metric	ripng metricout <i>value</i>	Optional 1 by default

Configuring RIPng Route Summarization

Follow these steps to configure RIPng route summarization:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--
Advertise a summary IPv6 prefix	ripng summary-address <i>ipv6-address</i> <i>prefix-length</i>	Required

Advertising a Default Route

Follow these steps to advertise a default route:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--
Advertise a default route	ripng default-route { only originate } [cost <i>cost</i>]	Required Not advertised by default



With this feature enabled, a default route is advertised via the specified interface regardless of whether the default route is available in the local IPv6 routing table.

Configuring a RIPng Route Filtering Policy

You can reference a configured IPv6 ACL or prefix list to filter received/advertised routing information as needed. For filtering outbound routes, you can also specify a routing protocol from which to filter routing information redistributed.

Follow these steps to configure a RIPng route filtering policy:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIPng view	ripng [<i>process-id</i>]	--
Configure a filter policy to filter incoming routes	filter-policy { <i>acl6-number</i> ipv6-prefix <i>ipv6-prefix-name</i> } import	Required By default, RIPng does not filter incoming routing information.
Configure a filter policy to filter outgoing routes	filter-policy { <i>acl6-number</i> ipv6-prefix <i>ipv6-prefix-name</i> } export [<i>protocol</i> [<i>process-id</i>]]	Required By default, RIPng does not filter outgoing routing information.

Configuring a Priority for RIPng

Any routing protocol has its own protocol priority used for optimal route selection. You can set a priority for RIPng manually. The smaller the value is, the higher the priority is.

Follow these steps to configure a RIPng priority:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter RIPng view	ripng [<i>process-id</i>]	-
Configure a RIPng priority	preference [route-policy <i>route-policy-name</i>] <i>preference</i>	Optional By default, the RIPng priority is 100.

Configuring RIPng Route Redistribution

Follow these steps to configure RIPng route redistribution:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIPng view	ripng [<i>process-id</i>]	--
Configure a default routing metric for redistributed routes	default cost <i>cost</i>	Optional By default, the default metric of redistributed routes is 0.
Redistribute routes from another routing protocol	import-route <i>protocol</i> [<i>process-id</i>] [allow-ibgp] [cost <i>cost</i> route-policy <i>route-policy-name</i>] *	Required No route redistribution is configured by default.

Tuning and Optimizing the RIPng Network

This section describes how to tune and optimize the performance of the RIPng network as well as applications under special network environments. Before tuning and optimizing the RIPng network, complete the following tasks:

- Configure a network layer address for each interface
- Configure the basic RIPng functions

This section covers the following topics:

- “Configuring RIPng Timers” on page 436
- “Configuring Split Horizon and Poison Reverse” on page 437
- “Configuring Zero Field Check on RIPng Packets” on page 438
- “Configuring the Maximum Number of Equal Cost Routes for Load Balancing” on page 438

Configuring RIPng Timers

You can adjust RIPng timers to optimize the performance of the RIPng network.

Follow these steps to configure RIPng timers:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter RIPng view	ripng [<i>process-id</i>]	-

To do...	Use the command...	Remarks
Configure RIPng timers	timers { garbage-collect <i>garbage-collect-value</i> suppress <i>suppress-value</i> timeout <i>timeout-value</i> update <i>update-value</i> } *	Optional. The RIPng timers have the following defaults: <ul style="list-style-type: none"> ■ 30 seconds for the update timer ■ 180 seconds for the timeout timer ■ 120 seconds for the suppress timer ■ 120 seconds for the garbage-collect timer



When adjusting RIPng timers, you should consider the network performance and perform unified configurations on routers running RIPng to avoid unnecessary network traffic increase or route oscillation.

Configuring Split Horizon and Poison Reverse



If both the split horizon and poison reverse are configured, only the poison reverse function takes effect.

Configure the split horizon

The split horizon function disables a route learned from an interface from being advertised via the interface to prevent routing loops between neighbors.

Follow these steps to configure the split horizon:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--
Enable the split horizon function	ripng split-horizon	Optional Enabled by default



Generally, you are recommended to enable the split horizon to prevent routing loops.

Configuring the poison reverse function

The poison reverse function enables a route learned from an interface to be advertised via the interface. However, the metric of the route is set to 16. That is to say, the route is unreachable.

Follow these steps to configure poison reverse:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--

To do...	Use the command...	Remarks
Enable the poison reverse function	ripng poison-reverse	Required Disabled by default

Configuring Zero Field Check on RIPng Packets

Some fields in the RIPng packet must be zero. These fields are called zero fields. With zero field check on RIPng packets enabled, if such a field contains a non-zero value, the entire RIPng packet will be discarded. If you are sure that all packets are trustworthy, you can disable the zero field check to save the CPU processing time.

Follow these steps to configure RIPng zero field check:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIPng view	ripng [<i>process-id</i>]	--
Enable the zero field check	checkzero	Optional Enabled by default

Configuring the Maximum Number of Equal Cost Routes for Load Balancing

Follow these steps to configure the maximum number of equal cost RIPng routes for load balancing:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Enter RIPng view	ripng [<i>process-id</i>]	--
Configure the maximum number of equal cost RIPng routes for load balancing	maximum load-balancing <i>number</i>	Optional 4 by default

Displaying and Maintaining RIPng

To do...	Use the command...	Remarks
Display configuration information of a RIPng process	display ripng [<i>process-id</i>]	Available in any view
Display routes in the RIPng database	display ripng <i>process-id</i> database	Available in any view
Display the routing information of a specified RIPng process	display ripng <i>process-id</i> route	Available in any view
Display RIPng interface information	display ripng <i>process-id</i> interface [<i>interface-type</i> <i>interface-number</i>]	Available in any view

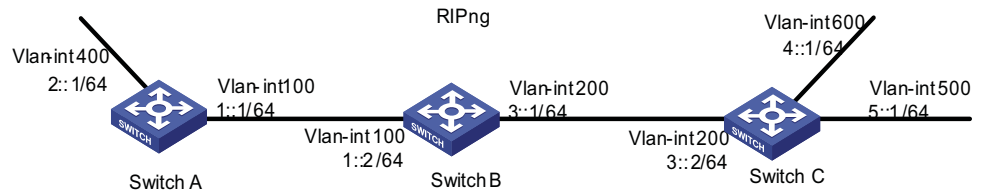
RIPng Configuration Example

Network requirements

As shown in Figure 142, all switches run RIPng. Configure Switch B to filter the route (3::/64) learnt from Switch C, which means the route will not be added to the routing table of Switch B, and Switch B will not forward it to Switch A.

Network diagram

Figure 142 Network diagram for RIPng configuration



Configuration procedure

- 1 Configure the IPv6 address for each interface (omitted)
- 2 Configure basic RIPng functions

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] ripng 1
[SwitchA-ripng-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ripng 1 enable
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 400
[SwitchA-Vlan-interface400] ripng 1 enable
[SwitchA-Vlan-interface400] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] ripng 1
[SwitchB-ripng-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ripng 1 enable
[SwitchB-Vlan-interface200] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ripng 1 enable
[SwitchB-Vlan-interface100] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ipv6
[SwitchC] ripng 1
[SwitchC-ripng-1] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] ripng 1 enable
[SwitchC-Vlan-interface200] quit
[SwitchC] interface Vlan-interface 500
[SwitchC-Vlan-interface500] ripng 1 enable
[SwitchC-Vlan-interface500] quit
[SwitchC] interface vlan-interface 600
[SwitchC-Vlan-interface600] ripng 1 enable
[SwitchC-Vlan-interface600] quit
```

Display the routing table of Switch B.

```
[SwitchB] display ripng 1 route
Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
-----

Peer FE80::20F:E2FF:FE23:82F5 on Vlan-interface100
Dest 1::/64,
    via FE80::20F:E2FF:FE23:82F5, cost 1, tag 0, A, 6 Sec
Dest 2::/64,
    via FE80::20F:E2FF:FE23:82F5, cost 1, tag 0, A, 6 Sec

Peer FE80::20F:E2FF:FE00:100 on Vlan-interface200
Dest 3::/64,
    via FE80::20F:E2FF:FE00:100, cost 1, tag 0, A, 11 Sec
Dest 4::/64,
    via FE80::20F:E2FF:FE00:100, cost 1, tag 0, A, 11 Sec
Dest 5::/64,
    via FE80::20F:E2FF:FE00:100, cost 1, tag 0, A, 11 Sec
```

Display the routing table of Switch A.

```
[SwitchA] display ripng 1 route
Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
-----

Peer FE80::200:2FF:FE64:8904 on Vlan-interface100
Dest 1::/64,
    via FE80::200:2FF:FE64:8904, cost 1, tag 0, A, 31 Sec
Dest 4::/64,
    via FE80::200:2FF:FE64:8904, cost 2, tag 0, A, 31 Sec
Dest 5::/64,
    via FE80::200:2FF:FE64:8904, cost 2, tag 0, A, 31 Sec
Dest 3::/64,
    via FE80::200:2FF:FE64:8904, cost 1, tag 0, A, 31 Sec
```

1 Configure Switch B to filter incoming and outgoing routes.

```
[SwitchB] acl ipv6 number 2000
[SwitchB-acl6-basic-2000] rule deny source 3::/64
[SwitchB-acl6-basic-2000] rule permit
[SwitchB-acl6-basic-2000] quit
[SwitchB] ripng 1
[SwitchB-ripng-1] filter-policy 2000 import
[SwitchB-ripng-1] filter-policy 2000 export
[SwitchB-ripng-1] quit
```

Display routing tables of Switch B and Switch A.

```
[SwitchB] display ripng 1 route
Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
-----

Peer FE80::20F:E2FF:FE23:82F5 on Vlan-interface100
Dest 1::/64,
    via FE80::20F:E2FF:FE23:82F5, cost 1, tag 0, A, 2 Sec
Dest 2::/64,
    via FE80::20F:E2FF:FE23:82F5, cost 1, tag 0, A, 2 Sec
```

```
Peer FE80::20F:E2FF:FE00:100 on Vlan-interface200
Dest 4::/64,
    via FE80::20F:E2FF:FE00:100, cost 1, tag 0, A, 5 Sec
Dest 5::/64,
    via FE80::20F:E2FF:FE00:100, cost 1, tag 0, A, 5 Sec
[SwitchA] display ripng 1 route
Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
-----
```

```
Peer FE80::20F:E2FF:FE00:1235 on Vlan-interface100
Dest 1::/64,
    via FE80::20F:E2FF:FE00:1235, cost 1, tag 0, A, 2 Sec
Dest 4::/64,
    via FE80::20F:E2FF:FE00:1235, cost 2, tag 0, A, 2 Sec
Dest 5::/64,
    via FE80::20F:E2FF:FE00:1235, cost 2, tag 0, A, 2 Sec
```


34

IPv6 OSPFv3 CONFIGURATION



- The term “router” in this document refers to a Layer 3 switch running routing protocols.
- The Switch 4800G only support single OSPFv3 process.

Introduction to OSPFv3

OSPFv3 Overview OSPFv3 is OSPF (Open Shortest Path First) version 3 for short, supporting IPv6 and compliant with RFC2740 (OSPF for IPv6).

Identical parts between OSPFv3 and OSPFv2:

- 32 bits router ID and area ID
- Packets: Hello, DD (Data Description), LSR (Link State Request), LSU (Link State Update), LSAck (Link State Acknowledgment)
- Mechanisms for finding neighbors and establishing adjacencies
- Mechanisms for LSA flooding and aging

Differences between OSPFv3 and OSPFv2:

- OSPFv3 now runs on a per-link basis, instead of on a per-IP-subnet basis.
- OSPFv3 supports multiple instances per link.
- OSPFv3 identifies neighbors by Router ID, while OSPFv2 by IP address.

OSPFv3 Packets OSPFv3 has also five types of packets: hello, DD, LSR, LSU, and LSAck.

The five packets have the same packet header, which different from the OSPFv2 packet header is only 16 bytes in length, has no authentication field, but is added with an Instance ID field to support multi-instance per link.

Figure 143 gives the OSPFv3 packet header.

Figure 143 OSPFv3 packet header

0	15	31
Version #	Type	Packet length
Router ID		
Area ID		
Checksum	Instance ID	0

Major fields:

- Version #: Version of OSPF, which is 3 for OSPFv3.
- Type: Type of OSPF packet, from 1 to 5 are hello, DD, LSR, LSU, and LSAck respectively.
- Packet Length: Packet length in bytes, including header.
- Instance ID: Instance ID for a link.
- 0: Reserved, which must be 0.

OSPFv3 LSA Types

OSPFv3 sends routing information in LSAs, which as defined in RFC2740 have the following types:

- Router-LSAs: Originated by all routers. This LSA describes the collected states of the router's interfaces to an area. Flooded throughout a single area only.
- Network-LSAs: Originated for broadcast and NBMA networks by the Designated Router. This LSA contains the list of routers connected to the network. Flooded throughout a single area only.
- Inter-Area-Prefix-LSAs: Similar to Type 3 LSA of OSPFv2, originated by ABRs (Area Border Routers), and flooded throughout the LSA's associated area. Each Inter-Area-Prefix-LSA describes a route with IPv6 address prefix to a destination outside the area, yet still inside the AS (an inter-area route).
- Inter-Area-Router-LSAs: Similar to Type 4 LSA of OSPFv2, originated by ABRs and flooded throughout the LSA's associated area. Each Inter-Area-Router-LSA describes a route to ASBR (Autonomous System Boundary Router).
- AS-external-LSAs: Originated by ASBRs, and flooded throughout the AS (except Stub and NSSA areas). Each AS-external-LSA describes a route to another Autonomous System. A default route can be described by an AS external LSA.
- Link-LSAs: A router originates a separate Link-LSA for each attached link. Link-LSAs have link-local flooding scope. Each Link-LSA describes the IPv6 address prefix of the link and Link-local address of the router.
- Intra-Area-Prefix-LSAs: Each Intra-Area-Prefix-LSA contains IPv6 prefix information on a router, stub area or transit area information, and has area flooding scope. It was introduced because Router-LSAs and Network-LSAs contain no address information now.

Timers of OSPFv3

Timers in OSPFv3 include:

- OSPFv3 packet timer
- LSA delay timer
- SPF timer

OSPFv3 packet timer

Hello packets are sent periodically between neighboring routers for finding and maintaining neighbor relationships, or for DR/BDR election. The hello interval must be identical on neighboring interfaces. The smaller the hello interval, the faster the network convergence speed and the bigger the network load.

If a router receives no hello packet from a neighbor after a period, it will declare the peer is down. The period is called dead interval.

After sending an LSA to its adjacency, a router waits for an acknowledgment from the adjacency. If no response is received after retransmission interval elapses, the router will send again the LSA. The retransmission interval must be longer than the round-trip time of the LSA in between.

LSA delay time

Each LSA has an age in the local LSDB (incremented by 1 per second), but an LSA is not aged on transmission. You need to add an LSA delay time into the age time before transmission, which is important for low speed networks.

SPF timer

Whenever LSDB changes, SPF recalculation happens. If recalculations become so frequent, a large amount of resources will be occupied, reducing operation efficiency of routers. You can adjust SPF calculation interval and delay time to protect networks from being overloaded due to frequent changes.

OSPFv3 Features Supported

- Basic features defined in RFC2740
- OSPFv3 stub area

Related RFCs

- RFC2740: OSPF for IPv6
- RFC2328: OSPF Version 2

IPv6 OSPFv3 Configuration Task List

Complete the following tasks to configure OSPFv3:

Task	Remarks
"Configuring OSPFv3 Basic Functions" on page 446	Required
"Configuring OSPFv3 Area Parameters" on page 446	"Configuring an OSPFv3 Stub Area" on page 447 "Configuring OSPFv3 Virtual Links" on page 447
"Configuring OSPFv3 Routing Information Management" on page 447	"Configuring OSPFv3 Route Summarization" on page 448
	"Configuring OSPFv3 Inbound Route Filtering" on page 448
	"Configuring Link Costs for OSPFv3 Interfaces" on page 448
	"Configuring the Maximum Number of OSPFv3 Load-balanced Routes" on page 449
	"Configuring a Priority for OSPFv3" on page 449
"Configuring OSPFv3 Route Redistribution" on page 449	Optional

Task	Remarks
"Tuning and Optimizing an OSPFv3 Network" on page 450	"Configuring OSPFv3 Timers" on page 450 Optional
	"Configuring the DR Priority for an Interface" on page 451 Optional
	"Ignoring MTU Check for DD Packets" on page 451 Optional
	"Disable Interfaces from Sending OSPFv3 Packets" on page 451 Optional
	"Enable the Logging on Neighbor State Changes" on page 451 Optional

Configuring OSPFv3 Basic Functions

- Prerequisites**
- Make neighboring nodes accessible with each other at network layer.
 - Enable IPv6 packet forwarding

Configuring OSPFv3 Basic Functions

Follow these steps to configure OSPFv3 basic functions:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable OSPFv3 and enter its view	ospfv3 [<i>process-id</i>]	Required
Specify a router ID	router-id <i>router-id</i>	Required
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable OSPFv3 on the interface	ospfv3 <i>process-id</i> area <i>area-id</i> [instance <i>instance-id</i>]	Required Not enabled by default



- *Configure an OSPFv3 process ID when enabling OSPFv3. The process ID takes effect locally, without affecting packet exchange between routers.*
- *When configuring a router ID, make sure each router has a unique ID.*
- *You need to specify a router ID manually, which is necessary to make OSPFv3 work.*

Configuring OSPFv3 Area Parameters

The stub area and virtual link support of OSPFv3 has the same principle and application environments with OSPFv2.

Splitting an OSPFv3 AS into multiple areas reduces the number of LSAs on networks and extends OSPFv3 application. For those non-backbone areas residing on the AS boundary, you can configure them as Stub areas to further reduce the size of routing tables on routers in these areas and the number of LSAs.

Non-backbone areas exchange routing information via the backbone area. Therefore, the backbone and non-backbone areas, including the backbone itself must maintain connectivity. In practice, necessary physical links may not be available for connectivity. You can configure virtual links to address it.

- Prerequisites**
- Enable IPv6 packet forwarding
 - Configure OSPFv3 basic functions

Configuring an OSPFv3 Stub Area

Follow these steps to configure an OSPFv3 stub area:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPFv3 view	ospfv3 [<i>process-id</i>]	-
Enter OSPFv3 area view	area <i>area-id</i>	-
Configure the area as a stub area	stub [no-summary]	Required Not configured by default
Configure the default route cost of sending a packet to the stub area	default-cost <i>value</i>	Optional Defaults to 1



- *Configurations on the OSPFv3 routers attached to the same area must be consistent. Otherwise, neighbor relationships cannot be established between adjacent routers.*
- *You cannot delete an OSPFv3 area directly. Only when you remove all configurations in area view and all interfaces attached to the area become down, can the area be removed automatically.*
- *All routers attached to a stub area must be configured with the **stub** command. The keyword **no-summary** is only available on the ABR.*
- *If you use the **stub** command with the keyword **no-summary** on an ABR, the ABR distributes a default summary LSA into the area rather than generating an AS-external-LSA or Inter-Area-Prefix-LSA. The stub area of this kind is also known as totally stub area.*

Configuring OSPFv3 Virtual Links

You can configure virtual links to maintain connectivity between non-backbone areas and the backbone, or in the backbone itself.

Follow these steps to configure a virtual link:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPFv3 view	ospfv3 [<i>process-id</i>]	-
Enter OSPFv3 area view	area <i>area-id</i>	-
Create and configure a virtual link	vlink-peer <i>router-id</i> [hello <i>seconds</i> retransmit <i>seconds</i> trans-delay <i>seconds</i> dead <i>seconds</i> instance <i>instance-id</i>] *	Required



*Both ends of a virtual link are ABRs that are configured with the **vlink-peer** command.*

Configuring OSPFv3 Routing Information Management

This section is to configure management of OSPF routing information advertisement and reception, and route redistribution from other protocols.

- Prerequisites**
- Enable IPv6 packet forwarding
 - Configure OSPFv3 basic functions

Configuring OSPFv3 Route Summarization

Follow these steps to configure route summarization between areas:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPFv3 view	ospfv3 [<i>process-id</i>]	-
Enter OSPFv3 area view	area <i>area-id</i>	-
Configure a summary route	abr-summary <i>ipv6-address</i> <i>prefix-length</i> [not-advertise]	Required Not configured by default



The **abr-summary** command is available on ABRs only. If contiguous network segments are available in an area, you can use the command to summarize them into one network segment on the ABR. The ABR will advertise only the summary route. Any LSA falling into the specified network segment will not be advertised, reducing the LSDB size in other areas.

Configuring OSPFv3 Inbound Route Filtering

You can configure OSPFv3 to filter routes that are computed from received LSAs according to some rules.

Follow these steps to configure inbound route filtering:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPFv3 view	ospfv3 [<i>process-id</i>]	-
Configure inbound route filtering	filter-policy { <i>acl-number</i> ipv6-prefix <i>ipv6-prefix-name</i> } import	Required Not configured by default



Use of the **filter-policy import** command can only filter routes computed by OSPFv3. Only routes not filtered can be added into the local routing table.

Configuring Link Costs for OSPFv3 Interfaces

You can configure OSPFv3 link costs for interfaces to adjust routing calculation.

Follow these steps to configure the link cost for an OSPFv3 interface:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure the cost for the interface	ospfv3 cost <i>value</i> [instance <i>instance-id</i>]	Optional 1 by default

Configuring the Maximum Number of OSPFv3 Load-balanced Routes

If multiple routes to a destination are available, using load balancing to send IPv6 packets on these routes in turn can improve link utility.

Follow these steps to configure the maximum number of load-balanced routes:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPFv3 view	ospfv3 [<i>process-id</i>]	-
Specify the maximum number of load-balanced routes	maximum load-balancing <i>maximum</i>	Optional 4 by default

Configuring a Priority for OSPFv3

A router may run multiple routing protocols. The system assigns a priority for each protocol. When these routing protocols find the same route, the route found by the protocol with the highest priority is selected.

Follow these steps to configure a priority for OSPFv3:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPFv3 view	ospfv3 [<i>process-id</i>]	-
Configure a priority for OSPFv3	preference [<i>ase</i>] [route-policy <i>route-policy-name</i>] <i>preference</i>	Optional By default, the priority of OSPFv3 interval routes is 10, and priority of OSPFv3 external routes is 150.

Configuring OSPFv3 Route Redistribution

Follow these steps to configure OSPFv3 route redistribution:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPFv3 view	ospfv3 [<i>process-id</i>]	-
Specify a default cost for redistributed routes	default cost <i>value</i>	Optional Defaults to 1
Redistribute routes from another protocol	import-route { isisv6 <i>process-id</i> ospfv3 <i>process-id</i> ripng <i>process-id</i> bgp4+ [allow-ibgp] direct static } [cost <i>value</i> type <i>type</i> route-policy <i>route-policy-name</i>] *	Required Not configured by default
Configure the filtering of outgoing redistributed routes	filter-policy { <i>acl6-number</i> ipv6-prefix <i>ipv6-prefix-name</i> } export [isisv6 <i>process-id</i> ospfv3 <i>process-id</i> ripng <i>process-id</i> bgp4+ direct static]	Optional Not configured by default



- Using the **import-route** command on a router makes the router become an ASBR.

- Since OSPFv3 is a link state based routing protocol, it cannot directly filter LSAs to be advertised. Therefore, you need to configure filtering redistributed routes before advertising routes that are not filtered in LSAs into the routing domain.
- Use of the **filter-policy export** command takes effect only on the local router. However, if the **import-route** command is not configured, executing the **filter-policy export** command does not take effect.

Tuning and Optimizing an OSPFv3 Network

This section describes configurations of OSPFv3 timers, interface DR priority, MTU check ignorance for DD packets, disabling interfaces from sending OSPFv3 packets.

OSPFv3 timers:

- Packet timer: Specified to adjust topology convergence speed and network load
- LSA delay timer: Specified especially for low speed links
- SPF timer: Specified to protect networks from being over consumed due to frequent network changes.

For a broadcast network, you can configure DR priorities for interfaces to affect DR/BDR election.

By disabling an interface from sending OSPFv3 packets, you can make other routers on the network obtain no information from the interface.

Prerequisites

- Enable IPv6 packet forwarding
- Configure OSPFv3 basic functions

Configuring OSPFv3 Timers

Follow these steps to configure OSPFv3 timers:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure the hello interval	ospfv3 timer hello <i>seconds</i> [instance <i>instance-id</i>]	Optional 10 seconds by default
Configure the dead interval	ospfv3 timer dead <i>seconds</i> [instance <i>instance-id</i>]	Optional 40 seconds by default
Configure the LSA retransmission interval	ospfv3 timer retransmit <i>interval</i> [instance <i>instance-id</i>]	Optional Defaults to 5 seconds
Configure the LSA transmission delay	ospfv3 trans-delay <i>seconds</i> [instance <i>instance-id</i>]	Optional Defaults to 1 second
Return to system view	quit	-
Enter OSPFv3 view	ospfv3 [<i>process-id</i>]	-
Configure the SPF timer	spf timers <i>delay-interval</i> <i>hold-interval</i>	Optional By default, <i>delay-interval</i> is 5 seconds, and <i>hold-interval</i> is 10 seconds



- The dead interval set on neighboring interfaces cannot be so short. Otherwise, a neighbor is easily considered down.
- The LSA retransmission interval cannot be so short; otherwise, unnecessary retransmissions occur.

Configuring the DR Priority for an Interface

Follow these steps to configure the DR priority for an interface:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type interface-number</i>	-
Configure the DR priority	ospfv3 dr-priority <i>priority</i> [instance <i>instance-id</i>]	Optional Defaults to 1



The DR priority of an interface determines the interface's qualification in DR election. Interfaces having the priority 0 cannot become a DR or BDR.

Ignoring MTU Check for DD Packets

When LSAs are few in DD packets, it is unnecessary to check MTU in DD packets in order to improve efficiency.

Follow these steps to ignore MTU check for DD packets:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type interface-number</i>	-
Ignore MTU check for DD packets	ospfv3 mtu-ignore [instance <i>instance-id</i>]	Required Not ignored by default

Disable Interfaces from Sending OSPFv3 Packets

Follow these steps to disable interfaces from sending OSPFv3 packets:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPFv3 view	ospfv3 [<i>process-id</i>]	-
Disable interfaces from sending OSPFv3 packets	silent-interface { <i>interface-type interface-number</i> all }	Required Not disabled by default



After an OSPF interface is set to silent, direct routes of the interface can still be advertised in Intra-Area-Prefix-LSAs via other interfaces, but other OSPFv3 packets cannot be advertised. Therefore, no neighboring relationship can be established on the interface. This feature can enhance the adaptability of OSPFv3 networking.

Enable the Logging on Neighbor State Changes

Follow these steps to enable the logging on neighbor state changes:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter OSPFv3 view	ospfv3 [<i>process-id</i>]	-

To do...	Use the command...	Remarks
Enable the logging on neighbor state changes	log-peer-change	Required Enabled by default

Displaying and Maintaining OSPFv3

To do...	Use the command...	Remarks
Display OSPFv3 debugging state information	display debugging ospfv3	Available in any view
Display OSPFv3 process brief information	display ospfv3 [<i>process-id</i>]	
Display OSPFv3 interface information	display ospfv3 interface [<i>interface-type interface-number</i> statistic]	
Display OSPFv3 LSDB information	display ospfv3 [<i>process-id</i>] lsdb [[external inter-prefix inter-router intra-prefix link network router] [<i>link-state-id</i>] [originate-router router-id] total]	
Display LSA statistics in OSPFv3 LSDB	display ospfv3 lsdb statistic	
Display OSPFv3 neighbor information	display ospfv3 [<i>process-id</i>] [area area-id] peer [[<i>interface-type interface-number</i>] [verbose] <i>peer-router-id</i>]	
Display OSPFv3 neighbor statistics	display ospfv3 peer statistic	
Display OSPFv3 routing table information	display ospfv3 [<i>process-id</i>] routing [<i>ipv6-address prefix-length</i> <i>ipv6-address/prefix-length</i> abr-routes asbr-routes all statistics]	
Display OSPFv3 area topology information	display ospfv3 [<i>process-id</i>] topology [area area-id]	
Display OSPFv3 virtual link information	display ospfv3 [<i>process-id</i>] vlink	
Display OSPFv3 next hop information	display ospfv3 [<i>process-id</i>] next-hop	
Display OSPFv3 link state request list information	display ospfv3 [<i>process-id</i>] request-list [{ external inter-prefix inter-router intra-prefix link network router } [<i>link-state-id</i>] [originate-router ip-address] statistics]	
Display OSPFv3 link state retransmission list information	display ospfv3 [<i>process-id</i>] retrans-list [{ external inter-prefix inter-router intra-prefix link network router } [<i>link-state-id</i>] [originate-router ip-address] statistics]	
Display OSPFv3 statistics	display ospfv3 statistic	

OSPFv3 Configuration Examples

Configuring OSPFv3 Areas

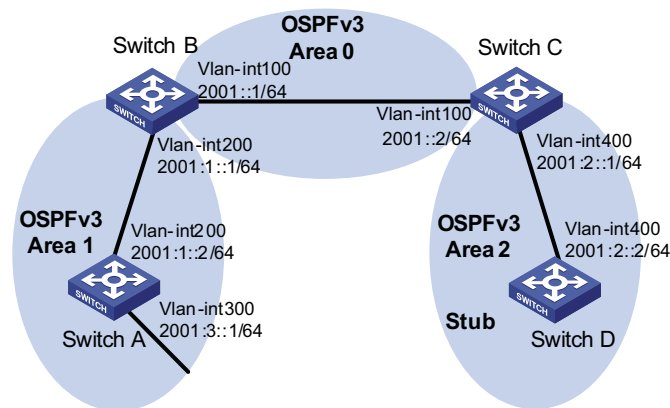
Network requirements

In the following figure, all switches run OSPFv3. The AS is split into three areas, in which, Switch B and Switch C act as ABRs to forward routing information between areas.

It is required to configure Area 2 as a stub area, reducing LSAs into the area without affecting route reachability.

Network diagram

Figure 144 Network diagram for OSPFv3 area configuration



Configuration procedure

- 1 Configure IPv6 addresses for interfaces (omitted)
- 2 Configure OSPFv3 basic functions

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] ospfv3
[SwitchA-ospfv3-1] router-id 1.1.1.1
[SwitchA-ospfv3-1] quit
[SwitchA] interface vlan-interface 300
[SwitchA-Vlan-interface300] ospfv3 1 area 1
[SwitchA-Vlan-interface300] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] ospfv3 1 area 1
[SwitchA-Vlan-interface200] quit
```

Configure Switch B

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] ospfv3
[SwitchB-ospf-1] router-id 2.2.2.2
[SwitchB-ospf-1] quit
[SwitchB] interface vlan-interface 100
```

```
[SwitchB-Vlan-interface100] ospfv3 1 area 0
[SwitchB-Vlan-interface100] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ospfv3 1 area 1
[SwitchB-Vlan-interface200] quit
```

Configure Switch C

```
<SwitchC> system-view
[SwitchC] ipv6
[SwitchC] ospfv3
[SwitchC-ospfv3-1] router-id 3.3.3.3
[SwitchC-ospfv3-1] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] ospfv3 1 area 0
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 400
[SwitchC-Vlan-interface400] ospfv3 1 area 2
[SwitchC-Vlan-interface400] quit
```

Configure Switch D

```
<SwitchD> system-view
[SwitchD] ipv6
[SwitchD] ospfv3
[SwitchD-ospfv3-1] router-id 4.4.4.4
[SwitchD-ospfv3-1] quit
[SwitchD] interface Vlan-interface 400
[SwitchD-Vlan-interface400] ospfv3 1 area 2
[SwitchD-Vlan-interface400] quit
```

Display OSPFv3 neighbor information on Switch B.

```
[SwitchB] display ospfv3 peer
```

```
OSPFv3 Area ID 0.0.0.0 (Process 1)
-----
Neighbor ID    Pri   State           Dead Time   Interface    Instance ID
3.3.3.3        1     Full/DR         00:00:39   Vlan100     0
```

```
OSPFv3 Area ID 0.0.0.1 (Process 1)
-----
Neighbor ID    Pri   State           Dead Time   Interface    Instance ID
1.1.1.1        1     Full/Backup     00:00:38   Vlan200     0
```

Display OSPFv3 neighbor information on Switch C.

```
[SwitchC] display ospfv3 peer
```

```
OSPFv3 Area ID 0.0.0.0 (Process 1)
-----
Neighbor ID    Pri   State           Dead Time   Interface    Instance ID
2.2.2.2        1     Full/Backup     00:00:39   Vlan100     0
```

```
OSPFv3 Area ID 0.0.0.2 (Process 1)
-----
Neighbor ID    Pri   State           Dead Time   Interface    Instance ID
4.4.4.4        1     Full/DR         00:00:38   Vlan400     0
```

Display OSPFv3 routing table information on Switch D.

```
[SwitchD] display ospfv3 routing
```

```
E1 - Type 1 external route,   IA - Inter area route,   I - Intra area route
E2 - Type 2 external route,   * - Seleted route
```

```

                                OSPFv3 Router with ID (4.4.4.4) (Process 1)
-----
*Destination: 2001::/64
  Type       : IA                      Cost       : 2
  NextHop    : FE80::F40D:0:93D0:1     Interface:  Vlan400

*Destination: 2001:1::/64
  Type       : IA                      Cost       : 3
  NextHop    : FE80::F40D:0:93D0:1     Interface:  Vlan400

*Destination: 2001:2::/64
  Type       : I                       Cost       : 1
  NextHop    : directly-connected      Interface:  Vlan400

*Destination: 2001:3::/64
  Type       : IA                      Cost       : 4
  NextHop    : FE80::F40D:0:93D0:1     Interface:  Vlan400

```

3 Configure Area 2 as a stub area

Configure Switch D

```
[SwitchD] ospfv3
[SwitchD-ospfv3-1] area 2
[SwitchD-ospfv3-1-area-0.0.0.2] stub
```

Configure Switch C, and specify the cost of the default route sent to the stub area as 10.

```
[SwitchC] ospfv3
[SwitchC-ospfv3-1] area 2
[SwitchC-ospfv3-1-area-0.0.0.2] stub
[SwitchC-ospfv3-1-area-0.0.0.2] default-cost 10
```

Display OSPFv3 routing table information on Switch D. You can find a default route is added, whose cost is the cost of the directly connected route plus the configured cost.

```
[SwitchD] display ospfv3 routing
E1 - Type 1 external route,   IA - Inter area route,   I - Intra area route
E2 - Type 2 external route,   * - Seleted route

                                OSPFv3 Router with ID (4.4.4.4) (Process 1)
-----
*Destination: ::/0
  Type       : IA                      Cost       : 11
  NextHop    : FE80::F40D:0:93D0:1     Interface:  Vlan400

*Destination: 2001::/64
  Type       : IA                      Cost       : 2
  NextHop    : FE80::F40D:0:93D0:1     Interface:  Vlan400

*Destination: 2001:1::/64
  Type       : IA                      Cost       : 3
  NextHop    : FE80::F40D:0:93D0:1     Interface:  Vlan400

*Destination: 2001:2::/64
  Type       : I                       Cost       : 1
  NextHop    : directly-connected      Interface:  Vlan400

*Destination: 2001:3::/64

```

```
Type      : IA
NextHop   : FE80::F40D:0:93D0:1
Cost      : 4
Interface : Vlan400
```

4 Configure Area 2 as a totally stub area

Configure Switch C, the ABR, to make Area 2 as a totally stub area.

```
[SwitchC-ospfv3-1-area-0.0.0.2] stub no-summary
```

Display OSPFv3 routing table information on Switch D. You can find route entries are reduced. All non direct routes are removed except the default route.

```
[SwitchD] display ospfv3 routing
E1 - Type 1 external route,   IA - Inter area route,   I - Intra area route
E2 - Type 2 external route,   * - Selected route

                OSPFv3 Router with ID (4.4.4.4) (Process 1)
-----
*Destination: ::/0
Type      : IA
NextHop   : FE80::F40D:0:93D0:1
Cost      : 11
Interface : Vlan400

*Destination: 2001:2::/64
Type      : I
NextHop   : directly-connected
Cost      : 1
Interface : Vlan400
```

Configuring OSPFv3 DR Election

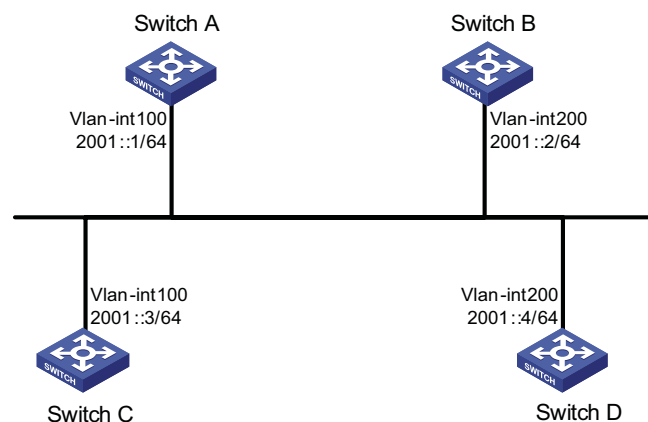
Network requirements

In the following figure:

- The priority of Switch A is 100, the highest priority on the network, so it will be the DR.
- The priority of Switch C is 2, the second highest priority on the network, so it will be the BDR.
- The priority of Switch B is 0, so it cannot become the DR.
- RouterD has the default priority 1.

Network diagram

Figure 145 Network diagram for OSPFv3 DR election configuration



Configuration procedure

- 1 Configure IPv6 addresses for interfaces (omitted)
- 2 Configure OSPFv3 basic functions

Configure Switch A

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] ospfv3
[SwitchA-ospfv3-1] router-id 1.1.1.1
[SwitchA-ospfv3-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ospfv3 1 area 0
[SwitchA-Vlan-interface100] quit
```

Configure Switch B

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] ospfv3
[SwitchB-ospfv3-1] router-id 2.2.2.2
[SwitchB-ospfv3-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ospfv3 1 area 0
[SwitchB-Vlan-interface200] quit
```

Configure Switch C

```
<SwitchC> system-view
[SwitchC] ipv6
[SwitchC] ospfv3
[SwitchC-ospfv3-1] router-id 3.3.3.3
[SwitchC-ospfv3-1] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] ospfv3 1 area 0
[SwitchC-Vlan-interface100] quit
```

Configure Switch D

```
<SwitchD> system-view
[SwitchD] ipv6
[SwitchD] ospfv3
[SwitchD-ospfv3-1] router-id 4.4.4.4
[SwitchD-ospfv3-1] quit
[SwitchD] interface vlan-interface 200
[SwitchD-Vlan-interface200] ospfv3 1 area 0
[SwitchD-Vlan-interface200] quit
```

Display neighbor information on Switch A. You can find the switches have the same default DR priority 1. In this case, the switch with the highest Router ID is elected as the DR. Therefore, Switch D is the DR, and Switch C is the BDR.

```
[SwitchA] display ospfv3 peer
                OSPFv3 Area ID 0.0.0.0 (Process 1)
-----
Neighbor ID    Pri   State           Dead Time   Interface   Instance ID
2.2.2.2        1     2-Way/DROther   00:00:36   Vlan200    0
3.3.3.3        1     Full/Backup     00:00:35   Vlan100    0
4.4.4.4        1     Full/DR         00:00:33   Vlan200    0
```

Display neighbor information on Switch D. You can find the neighbor states between Switch D and other switches are all full.

```
[SwitchD] display ospfv3 peer
      OSPFv3 Area ID 0.0.0.0 (Process 1)
-----
Neighbor ID    Pri   State           Dead Time   Interface   Instance ID
1.1.1.1        1   Full/DROther    00:00:30   Vlan100     0
2.2.2.2        1   Full/DROther    00:00:37   Vlan200     0
3.3.3.3        1   Full/Backup     00:00:31   Vlan100     0
```

3 Configure DR priorities for interfaces.

Configure the DR priority of VLAN-interface 100 as 100 on Switch A.

```
[SwitchA] interface Vlan-interface 100
[SwitchA-Vlan-interface100] ospfv3 dr-priority 100
[SwitchA-Vlan-interface100] quit
```

Configure the DR priority of VLAN-interface 200 as 0 on Switch B.

```
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] ospfv3 dr-priority 0
[SwitchB-Vlan-interface200] quit
```

#Configure the DR priority of Switch C as 2.

```
[SwitchC] interface Vlan-interface 100
[SwitchC-Vlan-interface100] ospfv3 dr-priority 2
[SwitchC-Vlan-interface100] quit
```

Display neighbor information on Switch A. You can find DR priorities have been updated, but DR and BDR are not changed.

```
[SwitchA] display ospfv3 peer
      OSPFv3 Area ID 0.0.0.0 (Process 1)
-----
Neighbor ID    Pri   State           Dead Time   Interface   Instance ID
2.2.2.2        0   2-Way/DROther   00:00:38   Vlan200     0
3.3.3.3        2   Full/Backup     00:00:32   Vlan100     0
4.4.4.4        1   Full/DR         00:00:36   Vlan200     0
```

#Display neighbor information on Switch D. You can find Switch D is still the DR.

```
[SwitchD] display ospfv3 peer
      OSPFv3 Area ID 0.0.0.0 (Process 1)
-----
Neighbor ID    Pri   State           Dead Time   Interface   Instance ID
1.1.1.1        100  Full/DROther    00:00:33   Vlan100     0
2.2.2.2        0   Full/DROther    00:00:36   Vlan200     0
3.3.3.3        2   Full/Backup     00:00:40   Vlan100     0
```

4 Restart DR/BDR election

Use the **shutdown** and **undo shutdown** commands on interfaces to restart DR/BDR election (omitted).

Display neighbor information on Switch A. You can find Switch C becomes the BDR.

```
[SwitchA] display ospfv3 peer
      OSPFv3 Area ID 0.0.0.0 (Process 1)
-----
Neighbor ID    Pri   State           Dead Time   Interface   Instance ID
2.2.2.2        0   Full/DROther    00:00:31   Vlan200     0
```

```
3.3.3.3      2      Full/Backup  00:00:39   Vlan100    0
4.4.4.4      1      Full/DROther 00:00:37   Vlan200    0
```

Display neighbor information on Switch D. You can find Switch A becomes the DR.

```
[SwitchD] display ospfv3 peer
          OSPFv3 Area ID 0.0.0.0 (Process 1)
-----
Neighbor ID  Pri  State           Dead Time  Interface  Instance ID
1.1.1.1     100 Full/DR         00:00:34   Vlan100    0
2.2.2.2      0   2-Way/DROther 00:00:34   Vlan200    0
3.3.3.3      2   Full/Backup    00:00:32   Vlan100    0
```

Troubleshooting OSPFv3 Configuration

No OSPFv3 Neighbor Relationship Established

Symptom

No OSPF neighbor relationship can be established.

Analysis

If the physical link and lower protocol work well, check OSPF parameters configured on interfaces. The two neighboring interfaces must have the same parameters, such as the area ID, network segment and mask, network type. If the network type is broadcast, at least one interface must have a DR priority higher than 0.

Process steps

- 1 Display neighbor information using the **display ospfv3 peer** command.
- 2 Display OSPFv3 interface information using the **display ospfv3 interface** command.
- 3 Ping the neighbor router's IP address to check connectivity.
- 4 Check OSPF timers. The dead interval on an interface must be at least four times the hello interval.
- 5 On a broadcast network, at least one interface must have a DR priority higher than 0.

Incorrect Routing Information

Symptom

OSPFv3 cannot find routes to other areas.

Analysis

The backbone area must maintain connectivity to all other areas. If a router connects to more than one area, at least one area must be connected to the backbone. The backbone cannot be configured as a Stub area.

In a Stub area, all routers cannot receive external routes, and all interfaces connected to the Stub area must be associated with the Stub area.

Solution

- 1 Use the **display ospfv3 peer** command to display OSPFv3 neighbors.
- 2 Use the **display ospfv3 interface** command to display OSPFv3 interface information.
- 3 Use the **display ospfv3 lsdb** command to display Link State Database information to check integrity.
- 4 Display information about area configuration using the **display current-configuration configuration** command. If more than two areas are configured, at least one area is connected to the backbone.
- 5 In a Stub area, all routers are configured with the **stub** command.
- 6 If a virtual link is configured, use the **display ospf vlink** command to check the neighbor state.



- IPv6 IS-IS supports all the features of IPv4 IS-IS except that it advertises IPv6 routing information instead. This document describes only IPv6 IS-IS exclusive configuration tasks. For other configuration tasks, refer to “IS-IS Configuration” on page 325.
- The term “router” in this document refers to a Layer 3 switch running routing protocols.

When configuring IPv6 IS-IS, go to these sections for information you are interested in:

- “Introduction to IPv6 IS-IS” on page 461
- “Configuring IPv6 IS-IS Basic Functions” on page 461
- “Configuring IPv6 IS-IS Routing Information Control” on page 462
- “Displaying and Maintaining IPv6 IS-IS” on page 463
- “IPv6 IS-IS Configuration Example” on page 464

Introduction to IPv6 IS-IS

The IS-IS routing protocol (Intermediate System-to-Intermediate System intra-domain routing information exchange protocol) supports multiple network protocols, including IPv6. IS-IS with IPv6 support is called IPv6 IS-IS dynamic routing protocol. The international switch fabric task force (IETF) defines two type-length-values (TLVs) and a new network layer protocol identifier (NLPID) to enable IPv6 support for IS-IS.

TLV is a variable field in the link state PDU or link state packet (LSP). The two TLVs are:

- IPv6 Reachability: Defines the prefix, metric of routing information to indicate the network reachability, with a type value of 236 (0xEC).
- IPv6 Interface Address: Similar with the “IP Interface Address” TLV of IPv4, it transforms the 32-bit IPv4 address to the 128-bit IPv6 address.

NLPID is an 8-bit field with a value of 142 (0x8E), which indicates the network layer protocol packet. If the IS-IS router supports IPv6, the advertised routing information must be marked with the NLPID.

Configuring IPv6 IS-IS Basic Functions



You can implement IPv6 inter-networking through configuring IPv6 IS-IS in IPv6 network environment.

Configuration Prerequisites

Before the configuration, accomplish the following tasks first:

- Enable IPv6 globally
- Configure IP addresses for interfaces, and make sure all neighboring nodes are reachable.
- Enable IS-IS

Configuration Procedure

Follow these steps to configure the basic functions of IPv6 IS-IS:

To do...	Use command to...	Remarks
Enter system view	system-view	--
Enable an IS-IS process and enter IS-IS view	isis [<i>process-id</i>]	Required Not enabled by default
Configure the network entity title for the IS-IS process	network-entity <i>net</i>	Required Not configured by default
Enable IPv6 for the IS-IS process	ipv6 enable	Required Disabled by default
Return to system view	quit	--
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	--
Enable IPv6 for an IS-IS process on the interface	isis ipv6 enable [<i>process-id</i>]	Required Disabled by default

Configuring IPv6 IS-IS Routing Information Control**Configuration Prerequisites**

You need to complete the IPv6 IS-IS basic function configuration before configuring this task.

Configuration Procedure

Follow these steps to configure IPv6 IS-IS routing information control:

To do...	Use command to...	Remarks
Enter system view	system-view	--
Enter IS-IS view	isis [<i>process-id</i>]	--
Define the priority for IPv6 IS-IS routes	ipv6 preference { route-policy <i>route-policy-name</i> <i>preference</i> } *	Optional 15 by default
Configure an IPv6 IS-IS summary route	ipv6 summary <i>ipv6-prefix</i> <i>prefix-length</i> [avoid-feedback] generate_null0_route [[level-1 level-1-2 level-2] tag <i>tag</i>] *	Optional Not configured by default
Generate an IPv6 IS-IS default route	ipv6 default-route-advertise [[level-1 level-2 level-1-2] route-policy <i>route-policy-name</i>]*	Optional No IPv6 default route is defined by default.
Configure IPv6 IS-IS to filter incoming routes	ipv6 filter-policy { <i>acl6-number</i> ipv6-prefix <i>ipv6-prefix-name</i> route-policy <i>route-policy-name</i> } import	Optional No filtering policy is defined by default

To do...	Use command to...	Remarks
Configure IPv6 IS-IS to redistribute routes from another routing protocol	ipv6 import-route <i>protocol</i> [<i>process-id</i>] [allow-ibgp] [cost cost-value] [level-1 level-2 level-1-2] route-policy <i>route-policy-name</i> tag <i>tag-value</i>] *	Optional Not configured by default
Configure the filtering of outgoing redistributed routes	ipv6 filter-policy { <i>acl6-number</i> ipv6-prefix <i>ipv6-prefix-name</i> route-policy <i>route-policy-name</i> } export [<i>protocol</i> [<i>process-id</i>]]	Optional Not configured by default
Enable route leaking	ipv6 import-route isisv6 level-2 into level-1 [filter-policy { <i>acl6-number</i> ipv6-prefix <i>ipv6-prefix-name</i> route-policy <i>route-policy-name</i> }] tag <i>tag</i>]*	Optional Not enabled by default
Specify the maximum number of equal-cost load balanced routes	ipv6 maximum load-balancing <i>number</i>	Optional 4 by default



The **ipv6 filter-policy export** command, usually used in combination with the **ipv6 import-route** command, filters redistributed routes when advertising them to other routers. If no protocol is specified, routes redistributed from all routing protocols are filtered before advertisement. If a protocol is specified, only routes redistributed from the routing protocol are filtered for advertisement.

Displaying and Maintaining IPv6 IS-IS

To do...	Use the command...	Remarks
Display brief IPv6 IS-IS information	display isis brief	Available in any view
Display the status of the debug switches	display isis debug-switches <i>process-id</i>	Available in any view
Display IS-IS enabled interface information	display isis interface [verbose] <i>process-id</i>	Available in any view
Display IS-IS license information	display isis license	Available in any view
Display LSDB information	display isis lsdb [[I1 I2 level-1 level-2]] [[lsp-id <i>lsp-id</i> lsp-name <i>lspname</i> local]] verbose] *] *	Available in any view
Display IS-IS mesh group information	display isis mesh-group [<i>process-id</i>]	Available in any view
Display the mapping table between the host name and system ID	display isis name-table [<i>process-id</i>]	Available in any view
Display IS-IS neighbor information	display isis peer [verbose] [<i>process-id</i>]	Available in any view
Display IPv6 IS-IS routing information	display isis route ipv6 [[level-1 level-2]] verbose] * [<i>process-id</i>]	Available in any view
Display SPF log information	display isis spf-log [<i>process-id</i>]	Available in any view
Display the statistics of the IS-IS process	display isis statistics [level-1 level-2 level-1-2] [<i>process-id</i>]	Available in any view
Clear all IS-IS data structure information	reset isis all [<i>process-id</i>]	Available in user view

To do...	Use the command...	Remarks
Clear the IS-IS data information of a neighbor	reset isis peer <i>system-id</i> [<i>process-id</i>]	Available in user view

IPv6 IS-IS Configuration Example

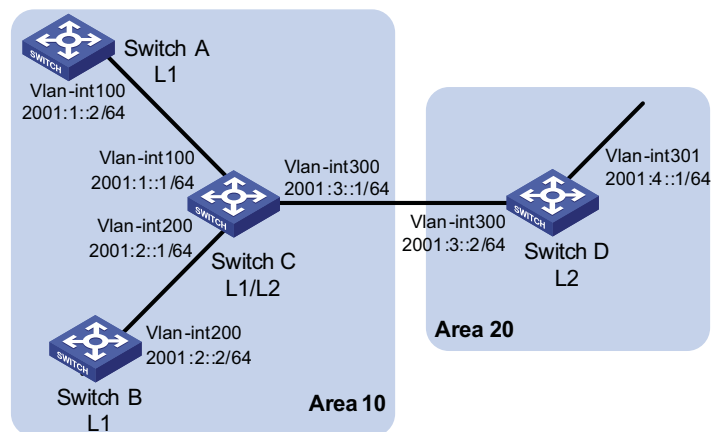
Network requirements

As shown in Figure 146, Switch A, Switch B, Switch C and Switch D reside in the same autonomous system, and all are enabled with IPv6.

Switch A and Switch B are Level-1 switches, Switch D is a Level-2 switch, and Switch C is a Level-1-2 switch. Switch A, Switch B, and Switch C are in area 10, while Switch D is in area 20.

Network diagram

Figure 146 Network diagram for IPv6 IS-IS basic configuration



Configuration procedure

- 1 Configure IPv6 addresses for interfaces (omitted)
- 2 Configure IPv6 IS-IS

Configure Switch A.

```
<SwitchA> system-view
[SwitchA] isis 1
[SwitchA-isis-1] is-level level-1
[SwitchA-isis-1] network-entity 10.0000.0000.0001.00
[SwitchA-isis-1] ipv6 enable
[SwitchA-isis-1] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] isis ipv6 enable 1
[SwitchA-Vlan-interface100] quit
```

Configure Switch B.

```
<SwitchB> system-view
[SwitchB] isis 1
[SwitchB-isis-1] is-level level-1
[SwitchB-isis-1] network-entity 10.0000.0000.0002.00
[SwitchB-isis-1] ipv6 enable
```

```
[SwitchB-isis-1] quit
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] isis ipv6 enable 1
[SwitchB-Vlan-interface200] quit
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] isis 1
[SwitchC-isis-1] network-entity 10.0000.0000.0003.00
[SwitchC-isis-1] ipv6 enable
[SwitchC-isis-1] quit
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] isis ipv6 enable 1
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] isis ipv6 enable 1
[SwitchC-Vlan-interface200] quit
[SwitchC] interface vlan-interface 300
[SwitchC-Vlan-interface300] isis ipv6 enable 1
[SwitchC-Vlan-interface300] quit
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] isis 1
[SwitchD-isis-1] is-level level-2
[SwitchD-isis-1] network-entity 20.0000.0000.0004.00
[SwitchD-isis-1] ipv6 enable
[SwitchD-isis-1] quit
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] isis ipv6 enable 1
[SwitchD-Vlan-interface300] quit
[SwitchD] interface vlan-interface 301
[SwitchD-Vlan-interface301] isis ipv6 enable 1
[SwitchD-Vlan-interface301] quit
```


36

IPv6 BGP CONFIGURATION



This chapter describes only configuration for IPv6 BGP. For other related information, refer to “BGP Configuration” on page 365.

When configuring IPv6 BGP, go to these sections for information you are interested in:

- “IPv6 BGP Overview” on page 467
- “Configuration Task List” on page 468
- “Configuring IPv6 BGP Basic Functions” on page 469
- “Controlling Route Distribution and Reception” on page 471
- “Configuring IPv6 BGP Route Attributes” on page 474
- “Tuning and Optimizing IPv6 BGP Networks” on page 476
- “Configuring a Large Scale IPv6 BGP Network” on page 478
- “Displaying and Maintaining IPv6 BGP Configuration” on page 482
- “IPv6 BGP Configuration Examples” on page 483
- “Troubleshooting IPv6 BGP Configuration” on page 486

IPv6 BGP Overview

BGP-4 manages only IPv4 routing information, thus other network layer protocols such as IPv6 are not supported.

To support multiple network layer protocols, IETF extended BGP-4 by introducing IPv6 BGP that is defined in RFC 2858 (multiprotocol extensions for BGP-4).

To implement IPv6 support, IPv6 BGP puts IPv6 network layer information into the attributes of network layer reachable information (NLRI) and NEXT_HOP.

NLRI attribute of IPv6 BGP involves:

- MP_REACH_NLRI: Multiprotocol Reachable NLRI, for advertisement of next hop information of reachable routes.
- MP_UNREACH_NLRI: Multiprotocol Unreachable NLRI, for withdrawal of unreachable routes.

The NEXT_HOP attribute of IPv6 BGP is identified by an IPv6 unicast address or IPv6 local link address.

IPv6 BGP utilizes BGP multiprotocol extensions for application in IPv6 networks. The original messaging and routing mechanisms of BGP are not changed.

Configuration Task List

Complete the following tasks to configure IPv6 BGP:

Task	Remarks	
"Configuring IPv6 BGP Basic Functions" on page 469	"Configuring an IPv6 Peer" on page 469	Required
	"Advertising a Local IPv6 Route" on page 469	Optional
	"Configuring a Preferred Value for Routes from a Peer/Peer Group" on page 469	Optional
	"Specifying the Source Interface for Establishing TCP Connections" on page 470	Optional
	"Allowing the establishment of a Non-Direct EBGp connection" on page 470	Optional
	"Configuring a Description for a Peer/Peer Group" on page 471	Optional
	"Disabling Session Establishment to a Peer/Peer Group" on page 471	Optional
"Controlling Route Distribution and Reception" on page 471	"Logging Peer State Changes" on page 471	Optional
	"Configuring IPv6 BGP Route Redistribution" on page 472	Optional
	"Advertising a Default Route to a Peer/Peer Group" on page 472	Optional
	"Configuring Route Distribution Policy" on page 472	Optional
	"Configuring Route Reception Policy" on page 473	Optional
	"Configuring IPv6 BGP and IGP Route Synchronization" on page 473	Optional
"Configuring IPv6 BGP Route Attributes" on page 474	"Configuring Route Dampening" on page 474	Optional
	"Configuring IPv6 BGP Preference and Default LOCAL_PREF and NEXT_HOP Attributes" on page 474	Optional
	"Configuring the MED Attribute" on page 475	Optional
"Tuning and Optimizing IPv6 BGP Networks" on page 476	"Configuring the AS_PATH Attribute" on page 475	Optional
	"Configuring IPv6 BGP Timers" on page 476	Optional
"Configuring a Large Scale IPv6 BGP Network" on page 478	"Configuring IPv6 BGP Soft Reset" on page 477	Optional
	"Configuring the Maximum Number of Load-Balanced Routes" on page 478	Optional
	"Configuring IPv6 BGP Peer Group" on page 478	Optional
	"Configuring IPv6 BGP Community" on page 480	Optional
	"Configuring an IPv6 BGP Route Reflector" on page 480	Optional

Configuring IPv6 BGP Basic Functions

Prerequisites Before configuring this task, you need to:

- Specify IP addresses for interfaces.
- Enable IPv6.



You need create a peer group before configuring basic functions for it. For related information, refer to “Configuring IPv6 BGP Peer Group” on page 478.

Configuring an IPv6 Peer Follow these steps to configure an IPv6 peer:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Specify a router ID	router-id <i>router-id</i>	Not enabled by default Optional
Enter IPv6 address family view	ipv6-family	Required if no IP addresses configured for Loopback interface and other interfaces
Specify an IPv6 peer and its AS number	peer <i>ipv6-address</i> as-number <i>as-number</i>	Optional Not configured by default

Advertising a Local IPv6 Route

Follow these steps to configure advertise a local route into the routing table:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Add a local route into IPv6 BGP routing table	network <i>ipv6-address prefix-length</i> [short-cut route-policy <i>route-policy-name</i>]	Required Not added by default

Configuring a Preferred Value for Routes from a Peer/Peer Group

Follow these steps to configure a preferred value for routes received from a peer/peer group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Configure a preferred value for routes received from a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } preferred-value <i>value</i>	Optional By default, the preferred value is 0.



*If you both reference a routing policy and use the command **peer** { *ipv6-group-name* | *ipv6-address* } **preferred-value** *value* to set a preferred value*

for routes from a peer, the routing policy sets a non-zero preferred value for routes matching it. Other routes not matching the routing policy uses the value set with the command. If the preferred value in the routing policy is zero, the routes matching it will also use the value set with the command. For information about using a routing policy to set a preferred value, refer to the **peer**

{ *ipv6-group-name* | *ipv6-address* } **route-policy** *route-policy-name* { **import** | **export** } command and the **apply preferred-value** *preferred-value* command.

Specifying the Source Interface for Establishing TCP Connections

Follow these steps to specify the source interface for establishing TCP connections to a BGP peer or peer group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Specify the source interface for establishing TCP connections to a BGP peer or peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } connect-interface <i>interface-type</i> <i>interface-number</i>	Required By default, IPv6 BGP uses the outbound interface of the best route to the BGP peer as the source interface for establishing a TCP connection.



- To improve stability and reliability, you can specify a loopback interface as the source interface for establishing TCP connections to a BGP peer. By doing so, a connection failure upon redundancy availability will not affect TCP connection establishment.
- To establish multiple BGP connections to a BGP router, you need to specify on the local router the respective source interfaces for establishing TCP connections to the peers on the peering BGP router; otherwise, the local BGP router may fail to establish TCP connections to the peers when using the outbound interfaces of the best routes as the source interfaces.

Allowing the establishment of a Non-Direct EBGP connection

Follow these steps to allow the establishment of EBGP connection to a non-directly connected peer/peer group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Allow the establishment of EBGP connection to a non directly connected peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } ebgp-max-hop [<i>hop-count</i>]	Required Not configured by default



CAUTION: In general, direct links should be available between EBGP peers. If not, you can use the **peer ebgp-max-hop** command to establish a multi-hop TCP connection in between. However, you need not use this command for direct EBGP connection with loopback interfaces.

Configuring a Description for a Peer/Peer Group

Follow these steps to configure description for a peer/peer group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Configure a description for a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } description <i>description-text</i>	Optional Not configured by default



The peer group to be configured with a description must have been created.

Disabling Session Establishment to a Peer/Peer Group

Follow these steps to disable session establishment to a peer/peer group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Disable session establishment to a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } ignore	Optional Not disabled by default

Logging Peer State Changes

Follow these steps to configure to log on the session and event information of a peer/peer group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enable logging of peer changes globally	log-peer-change	Optional Enabled by default
Enter IPv6 address family view	ipv6-family	-
Enable the state change logging for a peer or peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } log-change	Optional Enabled by default



*Refer to “Configuring BGP Basic Functions” on page 381 for information about the **log-peer-change** command.*

Controlling Route Distribution and Reception

The task includes routing information filtering, routing policy application and route dampening.

Prerequisites

Before configuring this task, you have:

- Enabled the IPv6 function
- Configured the IPv6 BGP basic functions

Configuring IPv6 BGP Route Redistribution

Follow these steps to configure IPv6 BGP route redistribution and filtering:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	-
Enter IPv6 address family view	ipv6-family	-
Enable default route redistribution into the IPv6 BGP routing table	default-route imported	Optional Not enabled by default
Enable route redistribution from another routing protocol	import-route <i>protocol</i> [<i>process-id</i>] [med <i>med-value</i> route-policy <i>route-policy-name</i>]*	Required Not enabled by default



If the **default-route imported** command is not configured, using the **import-route** command cannot redistribute any IGP default route.

Advertising a Default Route to a Peer/Peer Group

Follow these steps to configure to advertise default route to a peer/peer group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Advertise a default route to a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } default-route-advertise [route-policy <i>route-policy-name</i>]	Required Not advertised by default



With the **peer default-route-advertise** command used, the local router advertises a default route with itself as the next hop to the specified peer/peer group, regardless of whether the default route is available in the routing table.

Configuring Route Distribution Policy

Follow these steps to configure policies for route distribution:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Configure outbound route filtering	filter-policy { <i>acl6-number</i> ipv6-prefix <i>ipv6-prefix-name</i> } export [<i>protocol process-id</i>]	Required Not configured by default
Apply a routing policy to routes advertised to a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } route-policy <i>route-policy-name</i> export	Required Not applied by default
Specify an IPv6 ACL to filter routes advertised to a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } filter-policy <i>acl6-number</i> export	Required Not specified by default
Specify an AS path ACL to filter routes advertised to a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } as-path-acl <i>as-path-acl-number</i> export	Required Not specified by default

To do...	Use the command...	Remarks
Specify an IPv6 prefix list to filter routes advertised to a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } ipv6-prefix <i>ipv6-prefix-name</i> export	Required Not specified by default



- Members of a peer group must have the same outbound route policy with the peer group.
- IPv6 BGP advertises routes passing the specified policy to peers. Using the protocol argument can filter only the specified protocol routes. If no protocol specified, IPv6 BGP filters all routes to be advertised, including redistributed routes and routes imported using the **network** command.

Configuring Route Reception Policy

Follow these steps to configure route reception policy:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	-
Enter IPv6 address family view	ipv6-family	-
Configure inbound route filtering	filter-policy { <i>acl6-number</i> ipv6-prefix <i>ipv6-prefix-name</i> } import	Required Not configured by default
Apply a routing policy to routes from a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } route-policy <i>route-policy-name</i> import	Required Not applied by default
Specify an ACL to filter routes imported from a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } filter-policy <i>acl6-number</i> import	Required Not specified by default
Specify an AS path ACL to filter routing information imported from a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } as-path-acl <i>as-path-acl-number</i> import	Required Not specified by default
Specify an IPv6 prefix list to filter routing information imported from a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } ipv6-prefix <i>ipv6-prefix-name</i> import	Required Not specified by default
Specify the upper limit of address prefixes imported from a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } route-limit <i>limit</i> [<i>percentage</i>]	Optional By default, no limit on prefixes



- Only routes passing the specified policy can be added into the local IPv6 BGP routing table.
- Members of a peer group can have different inbound route policies.

Configuring IPv6 BGP and IGP Route Synchronization

With this feature enabled and when a non-BGP router is responsible for forwarding packets in an AS, IPv6 BGP speakers in the AS cannot advertise routing information to outside ASs unless all routers in the AS know the latest routing information.

By default, when a BGP router receives an IBGP route, it only checks the reachability of the route's next hop before advertisement. If the synchronization feature is configured, only the IBGP route is advertised by IGP can the route be advertised to EBGp peers.

Follow these steps to configure IPv6 BGP and IGP route synchronization:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Enable route synchronization between IPv6 BGP and IGP	synchronization	Required Not enabled by default

Configuring Route Dampening

Follow these steps to configure BGP route dampening:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Configure IPv6 BGP route dampening parameters	dampening [<i>half-life-reachable</i> <i>half-life-unreachable</i> <i>reuse suppress</i> <i>ceiling</i> route-policy <i>route-policy-name</i>]*	Optional Not configured by default

Configuring IPv6 BGP Route Attributes

This section describes how to use IPv6 BGP route attributes to modify BGP routing policy. These attributes are:

- IPv6 BGP protocol preference
- Default LOCAL_PREF attribute
- MED attribute
- NEXT_HOP attribute
- AS_PATH attribute

Prerequisites

Before configuring this task, you have:

- Enabled IPv6 function
- Configured IPv6 BGP basic functions

Configuring IPv6 BGP Preference and Default LOCAL_PREF and NEXT_HOP Attributes

Follow these steps to perform this configuration:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Configure preference values for IPv6 BGP external, internal, local routes	preference { <i>external-preference</i> <i>internal-preference</i> <i>local-preference</i> route-policy <i>route-policy-name</i> }	Optional The default preference values of external, internal and local routes are 255, 255, 130 respectively

To do...	Use the command...	Remarks
Configure the default value for local preference	default local-preference <i>value</i>	Optional The <i>value</i> defaults to 100
Advertise routes to a peer/peer group with the local router as the next hop	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } next-hop-local	Required By default, the feature is available for routes advertised to the EBGP peer/peer group, but not available to the IBGP peer/peer group



- To make sure an IBGP peer can find the correct next hop, you can configure routes advertised to the peer to use the local router as the next hop. If BGP load balancing is configured, the local router specifies itself as the next hop of outbound routes to a peer/peer group regardless of whether the **peer next-hop-local** command is configured.
- In a "third party next hop" network, that is, the two EBGP peers reside in a common broadcast subnet, the router does not specify itself as the next hop for routes to the EBGP peer by default, unless the **peer next-hop-local** command is configured.

Configuring the MED Attribute

Follow these steps to configure the MED attribute:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Configure a default MED value	default med <i>med-value</i>	Optional Defaults to 0
Enable to compare MED values of routes from different EBGP peers	compare-different-as-med	Optional Not enabled by default
Prioritize MED values of routes from each AS	bestroute compare-med	Optional Not configured by default
Prioritize MED values of routes from confederation peers	bestroute med-confederation	Optional Not configured by default

Configuring the AS_PATH Attribute

Follow these steps to configure the AS_PATH attribute:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Allow the local AS number to appear in AS_PATH of routes from a peer/peer group and specify the repeat times	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } allow-as-loop [<i>number</i>]	Optional Not allowed by default

To do...	Use the command...	Remarks
Specify a fake AS number for a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } fake-as <i>as-number</i>	Optional Not specified by default
Neglect the AS_PATH attribute for best route selection	bestroute as-path-neglect	Optional Not neglected by default
Configure to carry only the public AS number in updates sent to a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } public-as-only	Optional By default, BGP updates carry private AS number
Substitute local AS number for the AS number of a peer/peer group indicated in the AS_PATH attribute	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } substitute-as	Optional Not substituted by default

Tuning and Optimizing IPv6 BGP Networks

This section describes configurations of IPv6 BGP timers, IPv6 BGP connection soft reset and the maximum number of load balanced routes.

■ IPv6 BGP timers

After establishing an IPv6 BGP connection, two routers send keepalive messages periodically to each other to keep the connection. If a router receives no keepalive message from the peer after the holdtime elapses, it tears down the connection.

When establishing an IPv6 BGP connection, the two parties compare their holdtime values, taking the shorter one as the common holdtime. If the holdtime is 0, neither keepalive message is sent, nor holdtime is checked.

■ IPv6 BGP connection soft reset

After modifying a route selection policy, you have to reset IPv6 BGP connections to make the new one take effect, causing a short time disconnection. The current IPv6 BGP implementation supports the route-refresh feature that enables dynamic IPv6 BGP routing table refresh without needing to disconnect IPv6 BGP links.

With this feature enabled on all IPv6 BGP routers in a network, when a routing policy modified on a router, the router advertises a route-refresh message to its peers, which then send their routing information back to the router. Therefore, the local router can perform dynamic routing information update and apply the new policy without tearing down connections.

If a router not supporting route-refresh exists in the network, you need to configure the **peer keep-all-routes** command on the router to save all route updates, and then use the **refresh bgp ipv6** command to soft-reset IPv6 BGP connections.

Prerequisites Before configuring IPv6 BGP timers, you have:

- Enabled IPv6 function
- Configured IPv6 BGP basic functions

Configuring IPv6 BGP Timers

Follow these steps to configure IPv6 BGP timers:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Configure IPv6 BGP timers	Specify keepalive interval and holdtime Configure keepalive interval and holdtime for a peer/peer group	Optional The keepalive interval defaults to 60 seconds, holdtime defaults to 180 seconds.
Configure the interval for sending the same update to a peer/peer group	timer <i>keepalive keepalive hold holdtime</i> peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } timer <i>keepalive keepalive hold holdtime</i> peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } route-update-interval <i>seconds</i>	Optional The interval for sending the same update to an IBGP peer or an EBGP peer defaults to 15 seconds or 30 seconds



- *Timers configured using the **timer** command have lower priority than timers configured using the **peer timer** command.*
- *The holdtime interval must be at least three times the keepalive interval.*

Configuring IPv6 BGP Soft Reset

Enable route refresh

Follow these steps to enable route refresh:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Enable route refresh	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } capability-advertise route-refresh	Optional Enabled by default

Perform manual soft-reset

Follow these steps to perform manual soft reset:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Save all routes from a peer/peer group, not letting them go through the inbound policy	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } keep-all-routes	Optional Not saved by default.
Return to user view	return	Required
Soft-reset BGP connections manually	refresh bgp ipv6 { <i>all</i> <i>ipv6-address</i> <i>group</i> <i>ipv6-group-name</i> <i>external</i> <i>internal</i> } { <i>export</i> <i>import</i> }	



If the **peer keep-all-routes** command is used, all routes from the peer/peer group will be saved regardless of whether the filtering policy is available. These routes will be used to generate IPv6 BGP routes after soft-reset is performed.

Configuring the Maximum Number of Load-Balanced Routes

Follow these steps to configure the maximum number of load balanced routes:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Configure the maximum number of load balanced routes	balance <i>number</i>	Required By default, no load balancing is enabled.

Configuring a Large Scale IPv6 BGP Network

In a large-scale IPv6 BGP network, configuration and maintenance become no convenient due to too many peers. In this case, configuring peer groups makes management easier and improves route distribution efficiency. Peer group includes IBGP peer group, where peers belong to the same AS, and EBGP peer group, where peers belong to different ASs. If peers in an EBGP group belong to the same external AS, the EBGP peer group is a pure EBGP peer group, and if not, a mixed EBGP peer group.

In a peer group, all members enjoy a common policy. Using the community attribute can make a set of IPv6 BGP routers in multiple ASs enjoy the same policy, because sending of community between IPv6 BGP peers is not limited by AS.

To guarantee connectivity between IBGP peers, you need to make them fully meshed, but it becomes unpractical when there are too many IBGP peers. Using route reflectors or confederation can solve it. In a large-scale AS, both of them can be used.

Confederation configuration of IPv6 BGP is identical to that of BGP4, so it is not mentioned here. The following describes:

- Configuring IPv6 BGP peer group
- Configuring IPv6 BGP community
- Configuring IPv6 BGP route reflector

Prerequisites

Before configuring IPv6 BGP peer group, you have:

- Made peer nodes accessible at network layer
- Enabled BGP and configured router ID.

Configuring IPv6 BGP Peer Group

Create an IBGP peer group

Follow these steps to create an IBGP group:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Enter BGP view	bgp <i>as-number</i>	Required Not enabled by default
Enter IPv6 address family view	ipv6-family	-
Create an IBGP peer group	group <i>ipv6-group-name</i> [internal]	Required
Add a peer into the group	peer <i>ipv6-address</i> group <i>ipv6-group-name</i> [as-number <i>as-number</i>]	Required Not added by default

Create a pure EBGP peer group

Follow these steps to configure a pure EBGP group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required Not enabled by default
Enter IPv6 address family view	ipv6-family	-
Create an EBGP peer group	group <i>ipv6-group-name</i> external	Required
Configure the AS number for the peer group	peer <i>ipv6-group-name</i> as-number <i>as-number</i>	Required Not configured by default
Add an IPv6 peer into the peer group	peer <i>ipv6-address</i> group <i>ipv6-group-name</i>	Required Not added by default



- *To create a pure EBGP peer group, you need to specify an AS number for the peer group.*
- *If a peer was added into an EBGP peer group, you cannot specify any AS number for the peer group.*

Create a mixed EBGP peer group

Follow these steps to create a mixed EBGP peer group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required Not enabled by default
Enter IPv6 address family view	ipv6-family	-
Create an EBGP peer group	group <i>ipv6-group-name</i> external	Required
Specify the AS number of an IPv6 peer	peer <i>ipv6-address</i> as-number <i>as-number</i>	Required Not specified by default
Add the IPv6 peer into the peer group	peer <i>ipv6-address</i> group <i>ipv6-group-name</i>	Required Not added by default



When creating a mixed EBGP peer group, you need to create a peer and specify its AS number that can be different from AS numbers of other peers, but you cannot specify AS number for the EBGP peer group.

Configuring IPv6 BGP Community

Advertise community attribute to a peer/peer group

Follow these steps to advertise community attribute to a peer/peer group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required Not enabled by default
Enter IPv6 address family view	ipv6-family	-
Advertise community attribute to a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } advertise-community	Required Not advertised by default
Advertise extended community attribute to a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } advertise-ext-community	Required Not advertised by default

Apply a routing policy to routes advertised to a peer/peer group

Follow these steps to apply a routing policy to routes advertised to a peer/peer group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Apply a routing policy to routes advertised to a peer/peer group	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } route-policy <i>route-policy-name</i> export	Required Not applied by default



When configuring IPv6 BGP community, you need to configure a routing policy to define the community attribute, and apply the routing policy to route advertisement.

Configuring an IPv6 BGP Route Reflector

Follow these steps to configure an IPv6 BGP route reflector:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter BGP view	bgp <i>as-number</i>	Required
Enter IPv6 address family view	ipv6-family	-
Configure the router as a route reflector and specify a peer/peer group as a client	peer { <i>ipv6-group-name</i> <i>ipv6-address</i> } reflect-client	Required Not configured by default
Enable route reflection between clients	reflect between-clients	Optional Enabled by default

To do...	Use the command...	Remarks
Configure the cluster ID of the route reflector	reflect cluster-id <i>cluster-id</i>	Optional By default, a route reflector uses its router ID as the cluster ID



- *In general, since the route reflector forwards routing information between clients, it is not required to make clients of a route reflector fully meshed. If clients are fully meshed, it is recommended to disable route reflection between clients to reduce routing costs.*
- *If a cluster has multiple route reflectors, you need to specify the same cluster ID for these route reflectors to avoid routing loops.*

Displaying and Maintaining IPv6 BGP Configuration

Displaying BGP

To do...	Use the command...	Remarks
Display IPv6 BGP peer group information	display bgp ipv6 group [<i>ipv6-group-name</i>]	Available in any view
Display IPv6 BGP advertised routing information	display bgp ipv6 network	
Display IPv6 BGP AS path information	display bgp ipv6 paths [<i>as-regular-expression</i>]	
Display IPv6 BGP peer/peer group information	display bgp ipv6 peer [<i>ipv6-group-name</i> log-info <i>ipv6-address</i> { log-info verbose }]	
Display IPv6 BGP routing table information	display bgp ipv6 routing-table [<i>ipv6-address prefix-length</i>]	
Display IPv6 BGP routing information matching an AS path ACL	display bgp ipv6 routing-table as-path-acl <i>as-path-acl-number</i>	
Display IPv6 BGP community routing information	display bgp ipv6 routing-table community [<i>aa.nn<1-13></i>] [no-advertise no-export no-export-subconfed]* [whole-match]	
Display IPv6 BGP routing information matching an IPv6 BGP community list	display bgp ipv6 routing-table community-list { <i>basic-community-list-number</i> whole-match } <i>adv-community-list-number</i> }&<1-16>	
Display dampened IPv6 BGP routing information	display bgp ipv6 routing-table dampened	
Display IPv6 BGP dampening parameter information	display bgp ipv6 routing-table dampening parameter	
Display IPv6 BGP routing information originated from different ASs	display bgp ipv6 routing-table different-origin-as	
Display IPv6 BGP routing flap statistics	display bgp ipv6 routing-table flap-info [regular-expression <i>as-regular-expression</i> as-path-acl <i>as-path-acl-number</i> <i>network-address</i> [<i>prefix-length</i> longer-match]]	
Display BGP routing information to or from an IPv6 peer	display bgp ipv6 routing-table peer <i>ipv6-address</i> { advertised-routes received-routes } [<i>network-address prefix-length</i> statistic]	
Display IPv6 BGP routing information matching a regular expression	display bgp ipv6 routing-table regular-expression <i>as-regular-expression</i>	
Display IPv6 BGP routing statistics	display bgp ipv6 routing-table statistic	

Resetting IPv6 BGP Connections

To do...	Use the command...	Remarks
Perform soft reset on IPv6 BGP connections	refresh bgp ipv6 { <i>ipv6-address</i> all external group <i>ipv6-group-name</i> internal } { export import }	Available in user view
Reset IPv6 BGP connections	reset bgp ipv6 { <i>as-number</i> <i>ipv6-address</i> [flap-info] } all group <i>ipv6-group-name</i> external internal }	

Clearing IPv6 BGP Information

To do...	Use the command...	Remarks
Clear dampened IPv6 BGP routing information and release suppressed routes	reset bgp ipv6 dampening [<i>ipv6-address prefix-length</i>]	Available in user view
Clear IPv6 BGP route flap information	reset bgp ipv6 flap-info [<i>ipv6-address/prefix-length</i> regex <i>as-path-regex</i> as-path-acl <i>as-path-acl-number</i>]	

IPv6 BGP Configuration Examples



Some examples for IPv6 BGP configuration are similar to those of BGP-4, so refer to "BGP Configuration" on page 365 for related information.

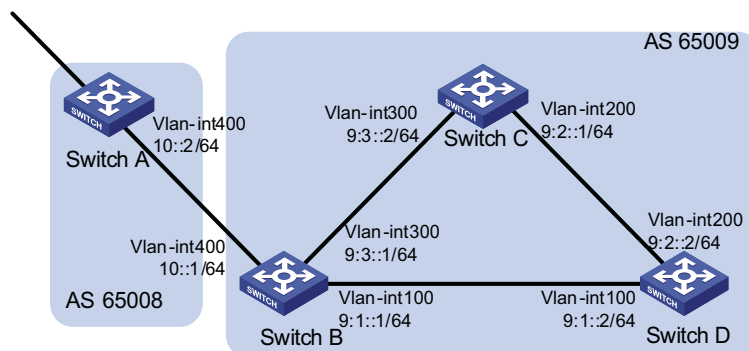
IPv6 BGP Basic Configuration

Network requirements

In the following figure are all IPv6 BGP switches. Between Switch A and Switch B is an EBGP connection. Switch B, Switch C and Switch D are IBGP fully meshed.

Network diagram

Figure 147 IPv6 BGP basic configuration network diagram



Configuration procedure

- 1 Configure IPv6 addresses for interfaces (omitted)
- 2 Configure IBGP connections

Configure Switch B.

```

<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] bgp 65009
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] ipv6-family
[SwitchB-bgp-af-ipv6] peer 9:1::2 as-number 65009
[SwitchB-bgp-af-ipv6] peer 9:3::2 as-number 65009
[SwitchB-bgp-af-ipv6] quit
[SwitchB-bgp] quit

```

Configure Switch C.

```

<SwitchC> system-view
[SwitchC] ipv6
[SwitchC] bgp 65009
[SwitchC-bgp] router-id 3.3.3.3
[SwitchC-bgp] ipv6-family
[SwitchC-bgp-af-ipv6] peer 9:3::1 as-number 65009
[SwitchC-bgp-af-ipv6] peer 9:2::2 as-number 65009
[SwitchC-bgp-af-ipv6] quit
[SwitchC-bgp] quit

```

Configure Switch D.

```

<SwitchD> system-view
[SwitchD] ipv6
[SwitchD] bgp 65009
[SwitchD-bgp] router-id 4.4.4.4
[SwitchD-bgp] ipv6-family
[SwitchD-bgp-af-ipv6] peer 9:1::1 as-number 65009
[SwitchD-bgp-af-ipv6] peer 9:2::1 as-number 65009
[SwitchD-bgp-af-ipv6] quit
[SwitchD-bgp] quit

```

3 Configure the EBGp connection

Configure Switch A.

```

<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] bgp 65008
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] ipv6-family
[SwitchA-bgp-af-ipv6] peer 10::1 as-number 65009
[SwitchA-bgp-af-ipv6] quit
[SwitchA-bgp] quit

```

Configure Switch B.

```

[SwitchB] bgp 65009
[SwitchB-bgp] ipv6-family
[SwitchB-bgp-af-ipv6] peer 10::2 as-number 65008

```

Display IPv6 peer information on Switch B.

```

[SwitchB] display bgp ipv6 peer

BGP local router ID : 2.2.2.2

```



```

Local AS number : 65009
Total number of peers : 3
Peers in established state : 3

Peer      V    AS  MsgRcvd  MsgSent  OutQ  PrefRcv  Up/Down  State
-----
10::2    4 65008      3        3     0        0 00:01:16 Established
9:3::2   4 65009      2        3     0        0 00:00:40 Established
9:1::2   4 65009      2        4     0        0 00:00:19 Established

```

Display IPv6 peer information on Switch C.

```

[SwitchC] display bgp ipv6 peer

BGP local router ID : 3.3.3.3
Local AS number : 65009
Total number of peers : 2
Peers in established state : 2

Peer      V    AS  MsgRcvd  MsgSent  OutQ  PrefRcv  Up/Down  State
-----
9:3::1    4 65009      4        4     0        0 00:02:18 Established
9:2::2    4 65009      4        5     0        0 00:01:52 Established

```

Switch A and B established an EBGP connection; Switch B, C and D established IBGP connections with each other.

IPv6 BGP Route Reflector Configuration

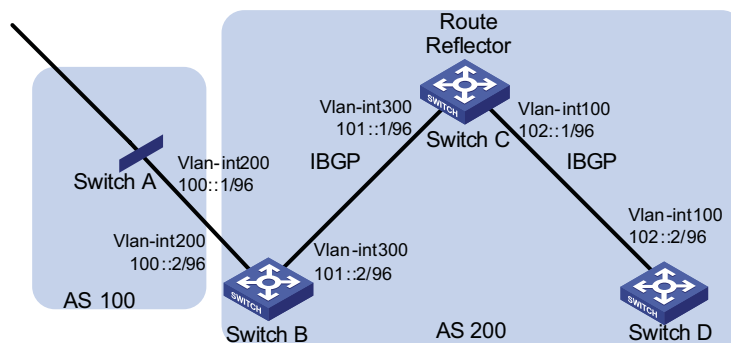
Network requirements

Switch B receives an EBGP update and sends it to Switch C, which is configured as a route reflector with two clients: Switch B and Switch D.

Switch B and Switch D need not establish an IBGP connection because Switch C reflects updates between them.

Network diagram

Figure 148 Network diagram for IPv6 BGP route reflector configuration



Configuration procedure

- 1 Configure IPv6 addresses for VLAN interfaces (omitted)
- 2 Configure IPv6 BGP basic functions

Configure Switch A.

```

<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] bgp 100

```

```
[SwitchA-bgp] router-id 1.1.1.1
[SwitchA-bgp] ipv6-family
[SwitchA-bgp-af-ipv6] peer 100::2 as-number 200
[SwitchA-bgp-af-ipv6] network 1:: 64
```

#Configure Switch B.

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] bgp 200
[SwitchB-bgp] router-id 2.2.2.2
[SwitchB-bgp] ipv6-family
[SwitchB-bgp-af-ipv6] peer 100::1 as-number 100
[SwitchB-bgp-af-ipv6] peer 101::1 as-number 200
[SwitchB-bgp-af-ipv6] peer 101::1 next-hop-local
```

Configure Switch C.

```
<SwitchC> system-view
[SwitchC] ipv6
[SwitchC] bgp 200
[SwitchC-bgp] router-id 3.3.3.3
[SwitchC-bgp] ipv6-family
[SwitchC-bgp-af-ipv6] peer 101::2 as-number 200
[SwitchC-bgp-af-ipv6] peer 102::2 as-number 200
```

Configure Switch D.

```
<SwitchD> system-view
[SwitchD] ipv6
[SwitchD] bgp 200
[SwitchD-bgp] router-id 4.4.4.4
[SwitchD-bgp] ipv6-family
[SwitchD-bgp-af-ipv6] peer 102::1 as-number 200
```

3 Configure route reflector

Configure Switch C as a route reflector, Switch B and Switch D as its clients.

```
[SwitchC-bgp-af-ipv6] peer 101::2 reflect-client
[SwitchC-bgp-af-ipv6] peer 102::2 reflect-client
```

Use the **display bgp ipv6 routing-table** command on Switch B and Switch D respectively, you can find both of them have learned the network 1::/64.

Troubleshooting IPv6 BGP Configuration

No IPv6 BGP Peer Relationship Established

Symptom

Display BGP peer information using the **display bgp ipv6 peer** command. The state of the connection to the peer cannot become established.

Analysis

To become IPv6 BGP peers, any two routers need to establish a TCP session using port 179 and exchange open messages successfully.

Processing steps

- 1** Use the **display current-configuration** command to verify the peer's AS number.
- 2** Use the **display bgp ipv6 peer** command to verify the peer's IPv6 address.
- 3** If the loopback interface is used, check whether the **peer connect-interface** command is configured.
- 4** If the peer is not directly connected, check whether the **peer ebgp-max-hop** command is configured.
- 5** Check whether a route to the peer is available in the routing table.
- 6** Use the **ping** command to check connectivity.
- 7** Use the **display tcp ipv6 status** command to check the TCP connection.
- 8** Check whether an ACL for disabling TCP port 179 is configured.

Introduction to Routing Policy

Routing Policy A routing policy is used on the router for route inspection, filtering, attributes modifying when routes are received, advertised, or redistributed.

When distributing or receiving routing information, a router can use a routing policy to filter routing information. For example, a router receives or advertises only routing information that matches the criteria of a routing policy; a routing protocol redistributes routes from another protocol only routes matching the criteria of a routing policy and modifies some attributes of these routes to satisfy its needs using the routing policy.

To implement a routing policy, you need to define a set of match criteria according to attributes in routing information, such as destination address, advertising router's address and so on. The match criteria can be set beforehand and then apply them to a routing policy for route distribution, reception and redistribution.

Filters Routing protocols can use six filters: ACL, IP prefix list, AS path ACL, community list, extended community list and routing policy.

ACL

When defining an ACL, you can specify IP addresses and prefixes to match destinations or next hops of routing information.

For ACL configuration, refer to "IPv6 ACL Configuration" on page 851.

IP prefix list

IP prefix list plays a role similar to ACL, but it is more flexible than ACL and easier to understand. When an IP prefix list is applied to filtering routing information, its matching object is the destination address of routing information.

An IP prefix list is identified by name. Each IP prefix list can comprise multiple items, and each item, which is identified by an index number, can specify a matching range in the network prefix format. The index number indicates the matching sequence of items in the IP prefix list.

During matching, the router compares the packet with the items in the ascending order. If one item is matched, the IP prefix list filter is passed, and the packet will not go to the next item.

AS-path

AS path is only applicable to IPv6 BGP. There is an AS-path field in the IPv6 BGP packet. An AS path list specifies matching conditions according to the AS-path field.

Community list

Community list only applies to IPv6 BGP. The IPv6 BGP packet contains a community attribute field to identify a community. A community list specifies matching conditions based on the community attribute.

Extended community list

Extended community list (extcommunity-list) applies to IPv6 BGP only. It is used for Route-Target extcommunity for VPN.

Routing policy

A routing policy is used to match against some attributes in given routing information and modify the attributes of the information if match conditions are satisfied. It can reference the above mentioned filters to define its own match criteria.

A routing policy can comprise multiple nodes, which are in logic OR relationship. Each node is a match unit, and the system compares each node to a packet in the order of node sequence number. Once a node is matched, the routing policy is passed and the packet will not go through the next node.

Each node comprises a set of **if-match** and **apply** clauses. The **if-match** clauses define the match criteria. The matching objects are some attributes of routing information. The different **if-match** clauses on a node is in logical AND relationship. Only when the matching conditions specified by all the **if-match** clauses on the node are satisfied, can routing information pass the node. The **apply** clauses specify the actions to be performed after the node is passed, concerning the attribute settings for routing information.

Routing Policy Application

A routing policy is applied in two ways:

- When redistributing routes from other routing protocols, a routing protocol accepts only routes passing the routing policy.
- When receiving or advertising routing information, a routing protocol uses the routing policy to filter routing information.

Defining Filtering Lists**Prerequisites**

Before configuring this task, you need to decide on:

- IP-prefix list name
- Matching address range
- Extcommunity list sequence number

Defining an IPv6 Prefix List

Identified by name, each IPv6 prefix list can comprise multiple items. Each item specifies a matching address range in the form of network prefix, which is identified by index number.

During matching, the system compares the route to each item in the ascending order of index number. If one item is matched, the route passes the IP-prefix list, without needing to match the next item.

Follow these steps to define an IPv6 prefix list:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Define an IPv6 prefix list	ip ipv6-prefix <i>ipv6-prefix-name</i> [index <i>index-number</i>] { deny permit } <i>ipv6-address</i> <i>prefix-length</i> [greater-equal <i>min-prefix-length</i>] [less-equal <i>max-prefix-length</i>]	Required Not defined by default



*If all items are set to the **deny** mode, no routes can pass the IPv6 prefix list. Therefore, you need to define the **permit :: 0 less-equal 128** item following multiple **deny** mode items to allow other IPv6 routing information to pass.*

For example, the following configuration filters routes 2000:1::/48, 2000:2::/48 and 2000:3::/48, but allows other routes to pass.

```
<Sysname> system-view
[Sysname] ip ipv6-prefix abc index 10 deny 2000:1:: 48
[Sysname] ip ipv6-prefix abc index 20 deny 2000:2:: 48
[Sysname] ip ipv6-prefix abc index 30 deny 2000:3:: 48
[Sysname] ip ipv6-prefix abc index 40 permit :: 0 less-equal 128
```

Defining an AS Path List

You can define multiple items for an AS path ACL that is identified by number. During matching, the relation between items is logical OR, that is, if the route matches one of these items, it passes the AS path ACL.

Follow these steps to define an AS path ACL:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Define an AS path ACL	ip as-path <i>as-path-number</i> { deny permit } <i>regular-expression</i>	Required Not defined by default

Defining a Community List

You can define multiple items for a community list that is identified by number. During matching, the relation between items is logic OR, that is, if routing information matches one of these items, it passes the community list.

Follow these steps to define a community list:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...		Use the command...	Remarks
Define a community list	Define a basic community list	ip community-list <i>basic-comm-list-num</i> { deny permit } [<i>community-number-list</i>] [internet no-advertise no-export no-export-subconfed] *	Required to define either; Not defined by default
	Define an advanced community list	ip community-list <i>adv-comm-list-num</i> { deny permit } <i>regular-expression</i>	

Defining an Extended Community List

You can define multiple items for an extended community list that is identified by number. During matching, the relation between items is logic OR, that is, if routing information matches one of these items, it passes the extended community list.

Follow these steps to define an extended community list:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Define an extended community list	ip extcommunity-list <i>ext-comm-list-number</i> { deny permit } { rt <i>route-target</i> } <1-16>	Required Not defined by default

Configuring a Routing Policy

A routing policy is used to filter routing information according to some attributes, and modify some attributes of the routing information that matches the routing policy. Match criteria can be configured using filters above mentioned.

A routing policy can comprise multiple nodes, each node contains:

- **if-match** clauses: Define the match criteria that routing information must satisfy. The matching objects are some attributes of routing information.
- **apply** clauses: Specify the actions performed after specified match criteria are satisfied, concerning attribute settings for passed routing information.

Prerequisites

Before configuring this task, you have completed:

- Filtering list configuration
- Routing protocol configuration

You also need to decide on:

- Name of the routing policy, node sequence numbers
- Match criteria
- Attributes to be modified

Creating a Routing Policy

Follow these steps to create a routing policy:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a routing policy and enter its view	route-policy <i>route-policy-name</i> { permit deny } node <i>node-number</i>	Required



- If a node has the **permit** keyword specified, routing information meeting the node's conditions will be handled using the **apply** clauses of this node, without needing to match against the next node. If routing information does not meet the node's conditions, it will go to the next node for a match.
- If a node is specified as **deny**, the **apply** clauses of the node will not be executed. When routing information matches all **if-match** clauses of the node, it can neither pass the node, nor go to the next node. If route information cannot match any **if-match** clause of the node, it will go to the next node for a match.
- When a routing policy is defined with more than one node, at least one node should be configured with the **permit** keyword. If the routing policy is used to filter routing information, routing information that does not meet any node's conditions cannot pass the routing policy. If all nodes of the routing policy are set using the **deny** keyword, no routing information can pass it.

Defining if-match Clauses for the Routing Policy

Follow these steps to define if-match clauses for a route-policy:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter routing policy view	route-policy <i>route-policy-name</i> { permit deny } node <i>node-number</i>	Required
Match IPv6 routes having the next hop or source specified in the ACL or IP prefix list	if-match ipv6 { address next-hop route-source } { acl <i>acl-number</i> prefix-list <i>ipv6-prefix-name</i> }	Optional Not configured by default
Match IPv6 BGP routes having AS path attributes specified in the AS path list (s)	if-match as-path <i>as-path-number</i> &<1-16>	Optional Not configured by default
Match IPv6 BGP routes having community attributes in the specified community list(s)	if-match community { <i>basic-community-list-number</i> [whole-match] <i>adv-community-list-number</i> }&<1-16>	Optional Not configured by default
Match routes having the specified cost	if-match cost <i>value</i>	Optional Not configured by default
Match BGP routes having extended attributes contained in the extended community list(s)	if-match extcommunity <i>ext-comm-list-number</i> &<1-16>	Optional Not configured by default
Match routes having specified outbound interface(s)	if-match interface { <i>interface-type</i> <i>interface-number</i> }&<1-16>	Optional Not configured by default

To do...	Use the command...	Remarks
Match routes having the specified route type	if-match route-type { internal external-type1 external-type2 external-type1or2 is-is-level-1 is-is-level-2 nssa-external-type1 nssa-external-type2 nssa-external-type1or2 } *	Optional Not configured by default
Match the routes having the specified tag value	if-match tag <i>value</i>	Optional Not configured by default



- The **if-match** clauses of a route-policy are in logic AND relationship, namely, routing information has to satisfy all **if-match** clauses before being executed with **apply** clauses.
- You can specify no or multiple **if-match** clauses for a routing policy. If no **if-match** clause is specified, and the routing policy is in permit mode, all routing information can pass the node; if in deny mode, no routing information can pass.

Defining apply Clauses for the Routing Policy

Follow these steps to define apply clauses for a route-policy:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a routing policy and enter its view	route-policy <i>route-policy-name</i> { permit deny } node <i>node-number</i>	Required Not created by default
Set AS_Path attribute for IPv6 BGP routes	apply as-path <i>as-number</i> <1-10> [replace]	Optional Not set by default
Specify a community list according to which to delete community attributes of IPv6 BGP routing information	apply comm-list <i>comm-list-number</i> delete	Optional Not configured by default
Set community attribute for IPv6 BGP routes	apply community { none additive { <i>community-number</i> <1-16> <i>aa:nn</i> <1-16> } internet no-export-subconfed no-export no-advertise } * [additive] }	Optional Not set by default
Set a cost for routes	apply cost [+ -] <i>value</i>	Optional Not set by default
Set a cost type for routes	apply cost-type { external internal type-1 type-2 }	Optional Not set by default
Set the extended community attribute for IPv6 BGP routes	apply extcommunity { rt { <i>as-number:nn</i> <i>ip-address:nn</i> } }<1-16> [additive] }	Optional Not set by default
Set a next hop for IPv6 routes	apply ipv6 next-hop <i>ipv6-address</i>	Optional Not set by default
Redistribute routes to a specified ISIS level	apply isis { level-1 level-1-2 level-2 }	Optional Not configured by default

To do...	Use the command...	Remarks
Set a local preference for IPv6 BGP routes	apply local-preference <i>preference</i>	Optional Not set by default
Set an origin attribute for IPv6 BGP routes	apply origin { igp egp <i>as-number</i> incomplete }	Optional Not set by default
Set a preference for the matched routing protocol	apply preference <i>preference</i>	Optional Not set by default
Set a preferred value for IPv6 BGP routes	apply preferred-value <i>preferred-value</i>	Optional Not set by default
Set a tag value for the routes	apply tag <i>value</i>	Optional Not set by default



The **apply ipv6 next-hop** commands do not apply to redistributed IPv6 routes respectively.

Displaying and Maintaining the Routing Policy

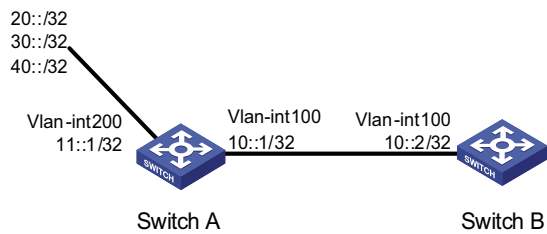
To do...	Use the command...	Remarks
Display IPv6 BGP AS path ACL information	display ip as-path [<i>as-path-number</i>]	Available in any view
Display IPv6 BGP community list information	display ip community-list [<i>basic-community-list-number</i> <i>adv-community-list-number</i>]	
Display IPv6 BGP extended community list information	display ip extcommunity-list [<i>ext-comm-list-number</i>]	
Display IPv6 prefix list statistics	display ip ipv6-prefix [<i>ipv6-prefix-name</i>]	
Display routing policy information	display route-policy [<i>route-policy-name</i>]	
Clear IPv6 prefix statistics	reset ip ipv6-prefix [<i>ipv6-prefix-name</i>]	

Routing Policy Configuration Example

Applying Routing Policy When Redistributing IPv6 Routes

Network requirements

- Enable RIPng on Switch A and Switch B.
- Configure three static routes on Switch A and apply a routing policy when redistributing static routes, making routes 20::0/32 and 40::0/32 pass, routes in 30::0/32 filtered out.
- Display RIPng routing table information on Switch B to verify the configuration.

Network diagram**Figure 149** Network diagram for routing policy application to route redistribution**Configuration procedure****1** Configure Switch A

Configure IPv6 addresses for VLAN-interface 100 and VLAN-interface 200.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ipv6 address 10::1 32
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] ipv6 address 11::1 32
[SwitchA-Vlan-interface200] quit
```

Enable RIPng on VLAN-interface 100.

```
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ripng 1 enable
[SwitchA-Vlan-interface100] quit
```

Configure three static routes.

```
[SwitchA] ipv6 route-static 20:: 32 11::2
[SwitchA] ipv6 route-static 30:: 32 11::2
[SwitchA] ipv6 route-static 40:: 32 11::2
```

Configure routing policy.

```
[SwitchA] ip ipv6-prefix a index 10 permit 30:: 32
[SwitchA] route-policy static2ripng deny node 0
[SwitchA-route-policy] if-match ipv6 address prefix-list a
[SwitchA-route-policy] quit
[SwitchA] route-policy static2ripng permit node 10
[SwitchA-route-policy] quit
```

Enable RIPng and redistribute static routes.

```
[SwitchA] ripng
[SwitchA-ripng-1] import-route static route-policy static2ripng
```

2 Configure Switch B.

Configure the IPv6 address for VLAN-interface 100.

```

[SwitchB] ipv6
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ipv6 address 10::2 32

# Enable RIPng on VLAN-interface 100.

[SwitchB-Vlan-interface100] ripng 1 enable
[SwitchB-Vlan-interface100] quit

# Enable RIPng.

[SwitchB] ripng

# Display RIPng routing table information.

[SwitchB-ripng-1] display ripng 1 route
Route Flags: A - Aging, S - Suppressed, G - Garbage-collect
-----

Peer FE80::7D58:0:CA03:1 on Vlan-interface 100
Dest 10::/32,
    via FE80::7D58:0:CA03:1, cost 1, tag 0, A, 18 Sec
Dest 20::/32,
    via FE80::7D58:0:CA03:1, cost 1, tag 0, A, 8 Sec
Dest 40::/32,
    via FE80::7D58:0:CA03:1, cost 1, tag 0, A, 3 Sec

```

Troubleshooting Routing Policy Configuration

IPv6 Routing Information Filtering Failure

Symptom

Filtering routing information failed, while routing protocol runs normally.

Analysis

At least one item of the IPv6 prefix list should be configured as permit mode, and at least one node of the Route-policy should be configured as permit mode.

Processing procedure

- 1 Use the **display ip ipv6-prefix** command to display IP prefix list information.
- 2 Use the **display route-policy** command to display routing policy information.

IPv6 BASICS CONFIGURATION

When configuring IPv6 basics, go to these sections for information you are interested in:

- "IPv6 Overview" on page 499
- "IPv6 Basics Configuration Task List" on page 508
- "Configuring Basic IPv6 Functions" on page 508
- "Configuring IPv6 NDP" on page 510
- "Configuring PMTU Discovery" on page 513
- "Configuring IPv6 TCP Properties" on page 514
- "Configuring ICMPv6 Packet Sending" on page 514
- "Configuring IPv6 DNS" on page 515
- "Displaying and Maintaining IPv6 Basics Configuration" on page 516
- "IPv6 Configuration Example" on page 517
- "Troubleshooting IPv6 Basics Configuration" on page 520



The term "router" or the router icon in this document refers to a router in a generic sense or a Layer 3 Ethernet switch running a routing protocol.

IPv6 Overview

Internet Protocol Version 6 (IPv6), also called IP next generation (IPng), was designed by the Internet Engineering Task Force (IETF) as the successor to Internet Protocol Version 4 (IPv4). The significant difference between IPv6 and IPv4 is that IPv6 increases the IP address size from 32 bits to 128 bits. This section covers the following:

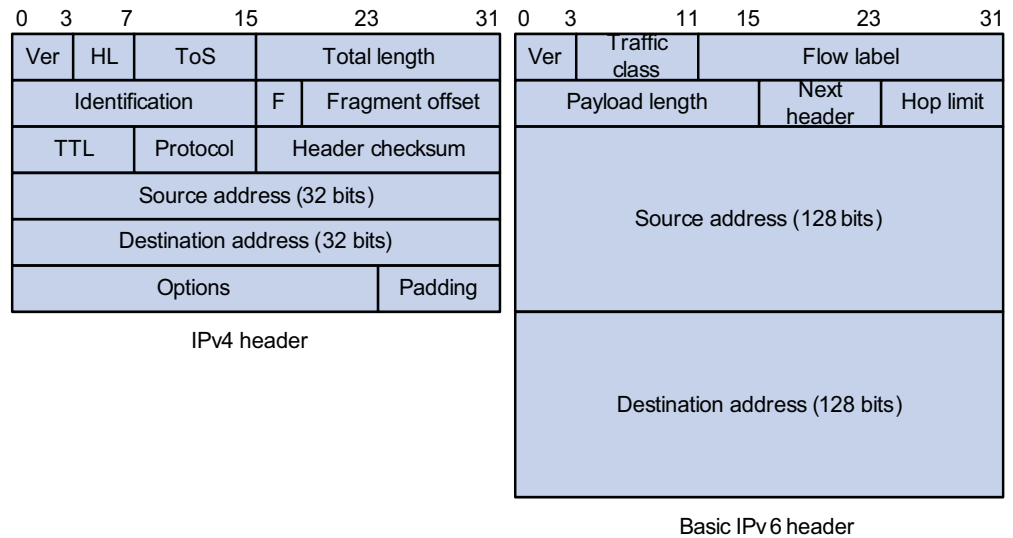
- "IPv6 Features" on page 499
- "Introduction to IPv6 Address" on page 501
- "Introduction to IPv6 Neighbor Discovery Protocol" on page 504
- "IPv6 PMTU Discovery" on page 507
- "Introduction to IPv6 DNS" on page 507
- "Protocols and Standards" on page 508

IPv6 Features **Header format simplification**

IPv6 cuts down some IPv4 header fields or move them to the IPv6 extension headers to reduce the length of the basic IPv6 header. IPv6 uses the basic header with a fixed length, thus making IPv6 packet handling simple and improving the forwarding efficiency. Although the IPv6 address size is four times that of IPv4

addresses, the size of basic IPv6 headers is 40 bytes and is only twice that of IPv4 headers (excluding the Options field).

Figure 150 Comparison between IPv4 packet header format and basic IPv6 packet header format



Adequate address space

The source and destination IPv6 addresses are both 128 bits (16 bytes) long. IPv6 can provide 3.4×10^{38} addresses to completely meet the requirements of hierarchical address division as well as allocation of public and private addresses.

Hierarchical address structure

IPv6 adopts the hierarchical address structure to quicken route search and reduce the system source occupied by the IPv6 routing table by means of route aggregation.

Automatic address configuration

To simplify the host configuration, IPv6 supports stateful and stateless address configuration.

- Stateful address configuration means that a host acquires an IPv6 address and related information from a server (for example, DHCP server).
- Stateless address configuration means that a host automatically configures an IPv6 address and related information on basis of its own link-layer address and the prefix information advertised by a router.

In addition, a host can generate a link-local address on basis of its own link-layer address and the default prefix (FE80::/64) to communicate with other hosts on the link.

Built-in security

IPv6 uses IPSec as its standard extension header to provide end-to-end security. This feature provides a standard for network security solutions and improves the interoperability between different IPv6 applications.

QoS support

The Flow Label field in the IPv6 header allows the device to label packets in a flow and provide special handling for these packets.

Enhanced neighbor discovery mechanism

The IPv6 neighbor discovery protocol is implemented through a group of Internet Control Message Protocol Version 6 (ICMPv6) messages that manages the information exchange between neighbor nodes on the same link. The group of ICMPv6 messages takes the place of Address Resolution Protocol (ARP) message, Internet Control Message Protocol version 4 (ICMPv4) router discovery message, and ICMPv4 redirection message to provide a series of other functions.

Flexible extension headers

IPv6 cancels the Options field in IPv4 packets but introduces multiple extension headers. In this way, IPv6 enhances the flexibility greatly to provide scalability for IP while improving the handling efficiency. The Options field in IPv4 packets contains 40 bytes at most, while the size of IPv6 extension headers is restricted by that of IPv6 packets.

Introduction to IPv6 Address

IPv6 address format

An IPv6 address is represented as a series of 16-bit hexadecimals, separated by colons. An IPv6 address is divided into eight groups, and the 16 bits of each group are represented by four hexadecimal numbers which are separated by colons, for example, 2001:0000:130F:0000:09C0:876A:130B.

To simplify the representation of IPv6 addresses, zeros in IPv6 addresses can be handled as follows:

- Leading zeros in each group can be removed. For example, the above-mentioned address can be represented in shorter format as 2001:0:130F:0:0:9C0:876A:130B.
- If an IPv6 address contains two or more consecutive groups of zeros, they can be replaced by the double-colon :: option. For example, the above-mentioned address can be represented in the shortest format as 2001:0:130F::9C0:876A:130B.



CAUTION: The double-colon :: option can be used only once in an IPv6 address. Otherwise, the device is unable to determine how many zeros double-colons represent when converting them to zeros to restore a 128-bit IPv6 address.

An IPv6 address consists of two parts: address prefix and interface ID. The address prefix and the interface ID are respectively equivalent to the network ID and the host ID in an IPv4 address.

An IPv6 address prefix is written in IPv6-address/prefix-length notation, where IPv6-address is an IPv6 address in any of the notations and prefix-length is a decimal number indicating how many bits from the utmost left of an IPv6 address are the address prefix.

IPv6 address classification

IPv6 addresses fall into three types: unicast address, multicast address, and anycast address.

- Unicast address: An identifier for a single interface, similar to an IPv4 unicast address. A packet sent to a unicast address is delivered to the interface identified by that address.
- Multicast address: An identifier for a set of interfaces (typically belonging to different nodes), similar to an IPv4 multicast address. A packet sent to a multicast address is delivered to all interfaces identified by that address.
- Anycast address: An identifier for a set of interfaces (typically belonging to different nodes). A packet sent to an anycast address is delivered to one of the interfaces identified by that address (the nearest one, according to the routing protocols' measure of distance).



There are no broadcast addresses in IPv6. Their function is superseded by multicast addresses.

The type of an IPv6 address is designated by the first several bits called format prefix. Table 47 lists the mappings between address types and format prefixes.

Table 47 Mapping between address types and format prefixes

Type	Format prefix (binary)	IPv6 prefix ID
Unicast address	Unassigned address	00...0 (128 bits)
	Loopback address	00...1 (128 bits)
	Link-local address	1111111010
	Site-local address	1111111011
	Global unicast address	other forms
Multicast address	11111111	
Anycast address	Anycast addresses are taken from unicast address space and are not syntactically distinguishable from unicast addresses.	

Unicast address

There are several forms of unicast address assignment in IPv6, including aggregatable global unicast address, link-local address, and site-local address.

- The aggregatable global unicast address, equivalent to an IPv4 public address, is provided for network service providers. The type of address allows efficient route prefix aggregation to restrict the number of global routing entries.
- The link-local address is used for communication between link-local nodes in neighbor discovery and stateless autoconfiguration. Routers must not forward any packets with link-local source or destination addresses to other links.
- IPv6 unicast site-local addresses are similar to private IPv4 addresses. Routers must not forward any packets with site-local source or destination addresses outside of the site (equivalent to a private network).
- Loopback address: The unicast address 0:0:0:0:0:0:0:1 (represented in the shortest format as ::1) is called the loopback address and may never be assigned to any physical interface. Like the loopback address in IPv4, it may be used by a node to send an IPv6 packet to itself.
- Unassigned address: The unicast address "::" is called the unassigned address and may not be assigned to any node. Before acquiring a valid IPv6 address, a

node may fill this address in the source address field of an IPv6 packet, but may not use it as a destination IPv6 address.

Multicast address

IPv6 multicast addresses listed in Table 48 are reserved for special purpose.

Table 48 Reserved IPv6 multicast addresses

Address	Application
FF01::1	Node-local scope all-nodes multicast address
FF02::1	Link-local scope all-nodes multicast address
FF01::2	Node-local scope all-routers multicast address
FF02::2	Link-local scope all-routers multicast address
FF05::2	Site-local scope all-routers multicast address

Besides, there is another type of multicast address: solicited-node address. A solicited-node multicast address is used to acquire the link-layer addresses of neighbor nodes on the same link and is also used for duplicate address detection (DAD). Each IPv6 unicast or anycast address has one corresponding solicited-node address. The format of a solicited-node multicast address is as follows:

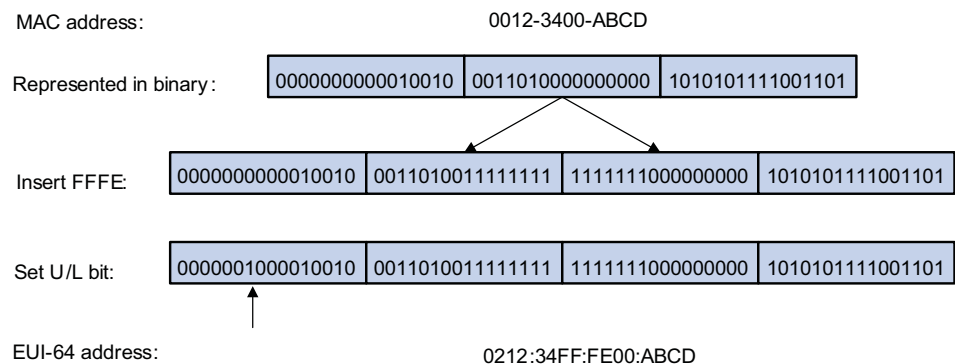
FF02:0:0:0:0:1:FFXX:XXXX

Where, FF02:0:0:0:0:1 FF is permanent and consists of 104 bits, and XX:XXXX is the last 24 bits of an IPv6 unicast or anycast address.

Interface identifier in IEEE EUI-64 format

Interface identifiers in IPv6 unicast addresses are used to identify interfaces on a link and they are required to be unique on that link. Interface identifiers in IPv6 unicast addresses are currently required to be 64 bits long. An interface identifier in IEEE EUI-64 format is derived from the link-layer address of that interface. Interface identifiers in IPv6 addresses are 64 bits long, while MAC addresses are 48 bits long. Therefore, the hexadecimal number FFFE needs to be inserted in the middle of MAC addresses (behind the 24 high-order bits). To ensure the interface identifier obtained from a MAC address is unique, it is necessary to set the universal/local (U/L) bit (the seventh high-order bit) to "1". Thus, an interface identifier in IEEE EUI-64 format is obtained.

Figure 151 Convert a MAC address into an EUI-64 interface identifier



Introduction to IPv6 Neighbor Discovery Protocol

IPv6 Neighbor Discovery Protocol (NDP) uses five types of ICMPv6 messages to implement the following functions:

- “Address resolution” on page 504
- “Neighbor reachability detection” on page 505
- “Duplicate address detection” on page 505
- “Router/prefix discovery and address autoconfiguration” on page 506
- “Redirection” on page 506

Table 49 lists the types and functions of ICMPv6 messages used by the NDP.

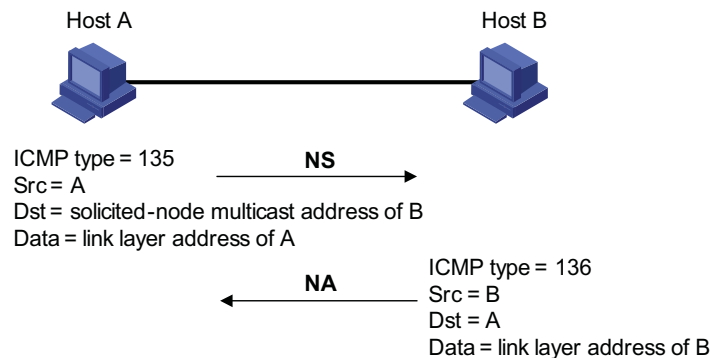
Table 49 Types and functions of ICMPv6 messages

ICMPv6 message	Number	Function
Neighbor solicitation (NS) message	135	Used to acquire the link-layer address of a neighbor Used to verify whether the neighbor is reachable Used to perform a duplicate address detection
Neighbor advertisement (NA) message	136	Used to respond to an NS message When the link layer changes, the local node initiates an NA message to notify neighbor nodes of the node information change.
Router solicitation (RS) message	133	After started, a node sends an RS message to request the router for an address prefix and other configuration information for the purpose of autoconfiguration.
Router advertisement (RA) message	134	Used to respond to an RS message With the RA message suppression disabled, the router regularly sends an RA message containing information such as prefix information options and flag bits.
Redirect message	137	When a certain condition is satisfied, the default gateway sends a redirect message to the source host so that the host can reselect a correct next hop router to forward packets.

The NDP mainly provides the following functions:

Address resolution

Similar to the ARP function in IPv4, a node acquires the link-layer addresses of neighbor nodes on the same link through NS and NA messages. Figure 152 shows how node A acquires the link-layer address of node B.

Figure 152 Address resolution

The address resolution procedure is as follows:

- 1 Node A multicasts an NS message. The source address of the NS message is the IPv6 address of an interface of node A and the destination address is the solicited-node multicast address of node B. The NS message contains the link-layer address of node A.
- 2 After receiving the NS message, node B judges whether the destination address of the packet corresponds to the solicited-node multicast address. If yes, node B can learn the link-layer address of node A, and unicasts an NA message containing its link-layer address.
- 3 Node A acquires the link-layer address of node B from the NA message.

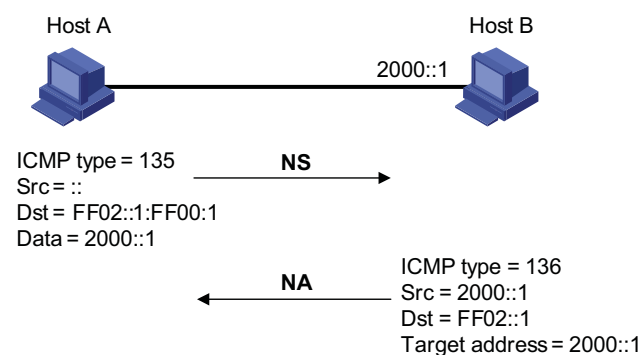
Neighbor reachability detection

After node A acquires the link-layer address of its neighbor node B, node A can verify whether node B is reachable according to NS and NA messages.

- 1 Node A sends an NS message whose destination address is the IPv6 address of node B.
- 2 If node A receives an NA message from node B, node A considers that node B is reachable. Otherwise, node B is unreachable.

Duplicate address detection

After node A acquires an IPv6 address, it will perform duplicate address detection (DAD) to determine whether the address is being used by other nodes (similar to the gratuitous ARP function of IPv4). DAD is accomplished through NS and NA messages. Figure 153 shows the DAD procedure.

Figure 153 Duplicate address detection

The DAD procedure is as follows:

- 1 Node A sends an NS message whose source address is the unassigned address :: and destination address is the corresponding solicited-node multicast address of the IPv6 address to be detected. The NS message contains the IPv6 address.
- 2 If node B uses this IPv6 address, node B returns an NA message. The NA message contains the IPv6 address of node B.
- 3 Node A learns that the IPv6 address is being used by node B after receiving the NA message from node B. Otherwise, node B is not using the IPv6 address and node A can use it.

Router/prefix discovery and address autoconfiguration

Router/prefix discovery means that a node locates the neighboring routers, and learns the prefix of the network where the host is located, and other configuration parameters from the received RA message.

Stateless address autoconfiguration means that a node automatically configures an IPv6 address according to the information obtained through router/prefix discovery.

The router/prefix discovery is implemented through RS and RA messages. The router/prefix discovery procedure is as follows:

- 1 After started, a node sends an RS message to request the router for the address prefix and other configuration information for the purpose of autoconfiguration.
- 2 The router returns an RA message containing information such as prefix information option. (The router also regularly sends an RA message.)
- 3 The node automatically configures an IPv6 address and other information for its interface according to the address prefix and other configuration parameters in the RA message.



- *In addition to an address prefix, the prefix information option also contains the preferred lifetime and valid lifetime of the address prefix. After receiving a periodic RA message, the node updates the preferred lifetime and valid lifetime of the address prefix accordingly.*
- *An automatically generated address is applicable within the valid lifetime and will be removed when the valid lifetime times out.*

Redirection

When a host is started, its routing table may contain only the default route to the gateway. When certain conditions are satisfied, the gateway sends an ICMPv6 redirect message to the source host so that the host can select a better next hop to forward packets (similar to the ICMP redirection function in IPv4).

The gateway will send an IPv6 ICMP redirect message when the following conditions are satisfied:

- The receiving interface is the forwarding interface.
- The selected route itself is not created or modified by an IPv6 ICMP redirect message.

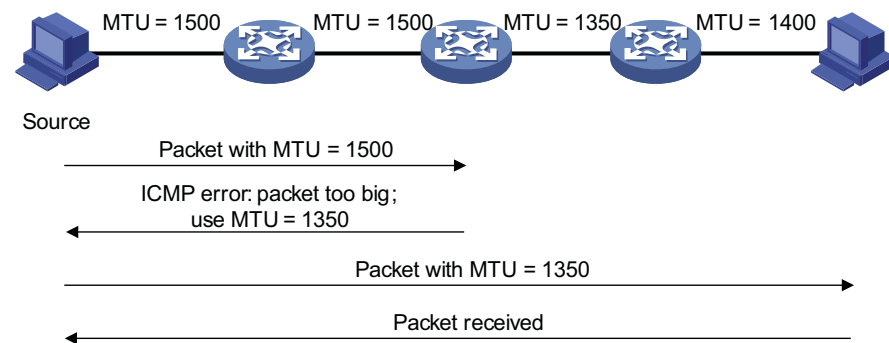
- The selected route is not the default route.
- The forwarded IPv6 packet does not contain any routing header.

IPv6 PMTU Discovery

The links that a packet passes from the source to the destination may have different MTUs. In IPv6, when the packet size exceeds the link MTU, the packet will be fragmented at the source end so as to reduce the processing pressure of the forwarding device and utilize network resources rationally.

The path MTU (PMTU) discovery mechanism is to find the minimum MTU of all links in the path from the source to the destination. Figure 154 shows the working procedure of the PMTU discovery.

Figure 154 Working procedure of the PMTU discovery



The working procedure of the PMTU discovery is as follows:

- 1 The source host uses its MTU to fragment packets and then sends them to the destination host.
- 2 If the MTU supported by the forwarding interface is less than the packet size, the forwarding device will discard the packet and return an ICMPv6 error packet containing the interface MTU to the source host.
- 3 After receiving the ICMPv6 error packet, the source host uses the returned MTU to fragment the packet again and then sends it.
- 4 Step 2 to step 3 are repeated until the destination host receives the packet. In this way, the minimum MTU of all links in the path from the source host to the destination host is determined.

Introduction to IPv6 DNS

In the IPv6 network, a Domain Name System (DNS) supporting IPv6 converts domain names into IPv6 addresses, instead of IPv4 addresses.

However, just like an IPv4 DNS, an IPv6 DNS also covers static domain name resolution and dynamic domain name resolution. The function and implementation of these two types of domain name resolution are the same as those of an IPv4 DNS. For details, refer to "DNS Overview" on page 971.

Usually, the DNS server connecting IPv4 and IPv6 networks not only contain A records (IPv4 addresses), but also AAAA records (IPv6 addresses). The DNS server can convert domain names into IPv4 addresses or IPv6 addresses. In this way, the DNS server implements the functions of both IPv6 DNS and IPv4 DNS.

Protocols and Standards

Protocols and standards related to IPv6 include:

- RFC 1881: IPv6 Address Allocation Management
- RFC 1887: An Architecture for IPv6 Unicast Address Allocation
- RFC 1981: Path MTU Discovery for IP version 6
- RFC 2375: IPv6 Multicast Address Assignments
- RFC 2460: Internet Protocol, Version 6 (IPv6) Specification.
- RFC 2461: Neighbor Discovery for IP Version 6 (IPv6)
- RFC 2462: IPv6 Stateless Address Autoconfiguration
- RFC 2463: Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification
- RFC 2464: Transmission of IPv6 Packets over Ethernet Networks
- RFC 2526: Reserved IPv6 Subnet Anycast Addresses
- RFC 3307: Allocation Guidelines for IPv6 Multicast Addresses
- RFC 3513: Internet Protocol Version 6 (IPv6) Addressing Architecture
- RFC 3596: DNS Extensions to Support IP Version 6

**IPv6 Basics
Configuration Task
List**

Complete the following tasks to perform IPv6 basics configuration:

Task	Remarks
"Configuring Basic IPv6 Functions" on page 508	Required
"Configuring IPv6 NDP" on page 510	Optional
"Configuring PMTU Discovery" on page 513	Optional
"Configuring IPv6 TCP Properties" on page 514	Optional
"Configuring ICMPv6 Packet Sending" on page 514	Optional
"Configuring IPv6 DNS" on page 515	Optional

**Configuring Basic IPv6
Functions**
**Enabling the IPv6 Packet
Forwarding Function**

Before IPv6-related configurations, you must enable the IPv6 packet forwarding function. Otherwise, an interface cannot forward IPv6 packets even if an IPv6 address is configured, resulting in communication failures in the IPv6 network.

Follow these steps to enable the IPv6 packet forwarding function:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the IPv6 packet forwarding function	ipv6	Required Disabled by default.

**Configuring an IPv6
Unicast Address**

IPv6 site-local addresses and aggregatable global unicast addresses can be configured in the following ways:

- EUI-64 format: When the EUI-64 format is adopted to form IPv6 addresses, the IPv6 address prefix of an interface is the configured prefix and the interface identifier is derived from the link-layer address of the interface.
- Manual configuration: IPv6 site-local addresses or aggregatable global unicast addresses are configured manually.

IPv6 link-local addresses can be configured in either of the following ways:

- Automatic generation: The device automatically generates a link-local address for an interface according to the link-local address prefix (FE80::/64) and the link-layer address of the interface.
- Manual assignment: IPv6 link-local addresses can be assigned manually.

Follow these steps to configure an IPv6 unicast address:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Enter interface view		interface <i>interface-type</i> <i>interface-number</i>	-
Configure an IPv6 aggregatable global unicast address or site-local address	Manually assign an IPv6 address	ipv6 address { <i>ipv6-address</i> <i>prefix-length</i> <i>ipv6-address/prefix-length</i> <i>h</i> }	Required to use either command. By default, no site-local address or aggregatable global unicast address is configured for an interface.
	Adopt the EUI-64 format to form an IPv6 address	ipv6 address <i>ipv6-address/prefix-length</i> <i>h</i> eui-64	
Configure an IPv6 link-local address	Automatically generate a link-local address	ipv6 address auto link-local	Optional By default, after an IPv6 site-local address or aggregatable global unicast address is configured for an interface, a link-local address will be generated automatically.
	Manually assign a link-local address for an interface	ipv6 address <i>ipv6-address</i> link-local	



- *After an IPv6 site-local address or aggregatable global unicast address is configured for an interface, a link-local address will be generated automatically. The automatically generated link-local address is the same as the one generated by using the **ipv6 address auto link-local** command. If a link-local address is manually assigned to an interface, this link-local address takes effect. If the manually assigned link-local address is removed, the automatically generated link-local address takes effect.*
- *The manual assignment takes precedence over the automatic generation. That is, if you first adopt the automatic generation and then the manual assignment, the manually assigned link-local address will overwrite the automatically generated one. If you first adopt the manual assignment and then the automatic generation, the automatically generated link-local address will not take effect and the link-local address of an interface is still the manually assigned one. If you delete the manually assigned address, the automatically generated link-local address is validated.*

- You need to execute the **ipv6 address auto link-local** command before the **undo ipv6 address auto link-local** command. However, if an IPv6 site-local address or aggregatable global unicast address is already configured for an interface, the interface still has a link-local address because the system automatically generates one for the interface. If no IPv6 site-local address or aggregatable global unicast address is configured, the interface has no link-local address.

Configuring IPv6 NDP

Configuring a Static Neighbor Entry

The IPv6 address of a neighbor node can be resolved into a link-layer address dynamically through NS and NA messages or through a manually configured neighbor entry.

The device uniquely identifies a static neighbor entry according to the IPv6 address and the layer 3 interface ID. Currently, there are two configuration methods:

- Configure an IPv6 address and link-layer address for a Layer 3 interface.
- Configure an IPv6 address and link-layer address for a port in a VLAN.

Follow these steps to configure a static neighbor entry:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure a static neighbor entry	ipv6 neighbor <i>ipv6-address mac-address</i> { <i>vlan-id port-type port-number</i> interface <i>interface-type interface-number</i> }	Required



CAUTION: You can adopt either of the two methods above to configure a static neighbor entry for a VLAN interface.

- After a static neighbor entry is configured by using the first method, the device needs to resolve the corresponding Layer 2 port information of the VLAN interface.
- If you adopt the second method to configure a static neighbor entry, you should ensure that the corresponding VLAN interface exists and that the layer 2 port specified by *port-type port-number* belongs to the VLAN specified by *vlan-id*. After a static neighbor entry is configured, the device relates the VLAN interface to an IPv6 address to uniquely identify a static neighbor entry.

Configuring the Maximum Number of Neighbors Dynamically Learned

The device can dynamically acquire the link-layer address of a neighbor node and add it into the neighbor table through NS and NA messages. Too large a neighbor table from which neighbor entries can be dynamically acquired may lead to the forwarding performance degradation of the device. Therefore, you can restrict the size of the neighbor table by setting the maximum number of neighbors that an interface can dynamically learn. When the number of dynamically learned neighbors reaches the threshold, the interface will stop learning neighbor information.

Follow these steps to configure the maximum number of neighbors dynamically learned:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure the maximum number of neighbors dynamically learned by an interface	ipv6 neighbors max-learning-num <i>number</i>	Optional

Configuring Parameters Related to an RA Message

You can configure whether the interface sends an RA message, the interval for sending RA messages, and parameters in RA messages. After receiving an RA message, a host can use these parameters to perform corresponding operations. Table 50 lists the configurable parameters in an RA message and their descriptions.

Table 50 Parameters in an RA message and their descriptions

Parameters	Description
Cur hop limit	When sending an IPv6 packet, a host uses the value of this parameter to fill the Cur Hop Limit field in IPv6 headers. Meanwhile, the value of this parameter is equal to the value of the Cur Hop Limit field in response messages of the device.
Prefix information options	After receiving the prefix information advertised by the device, the hosts on the same link can perform stateless autoconfiguration operations.
M flag	This field determines whether hosts use the stateful autoconfiguration to acquire IPv6 addresses. If the M flag is set to 1, hosts use the stateful autoconfiguration to acquire IPv6 addresses. Otherwise, hosts use the stateless autoconfiguration to acquire IPv6 addresses, that is, hosts configure IPv6 addresses according to their own link-layer addresses and the prefix information issued by the router.
O flag	This field determines whether hosts use the stateful autoconfiguration to acquire information other than IPv6 addresses. If the O flag is set to 1, hosts use the stateful autoconfiguration (for example, DHCP server) to acquire information other than IPv6 addresses. Otherwise, hosts use the stateless autoconfiguration to acquire information other than IPv6 addresses.
Router lifetime	This field is used to set the lifetime of the router that sends RA messages to serve as the default router of hosts. According to the router lifetime in the received RA messages, hosts determine whether the router sending RA messages can serve as the default router of hosts.
Retrans timer	If the device fails to receive a response message within the specified time after sending an NS message, the device will retransmit it.
Reachable time	After the neighbor reachability detection shows that a neighbor is reachable, the device considers the neighbor is reachable within the reachable time. If the device needs to send a packet to a neighbor after the reachable time expires, the device will again confirm whether the neighbor is reachable.



The values of the Retrans Timer field and the Reachable Time field configured for an interface are sent to hosts via RA messages. Furthermore, this interface sends NS messages at intervals of Retrans Timer and considers a neighbor reachable within the time of Reachable Time.

Follow these steps to configure parameters related to an RA message:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the current hop limit	ipv6 nd hop-limit <i>value</i>	Optional 64 by default.
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Disable the RA message suppression	undo ipv6 nd ra halt	Optional By default, RA messages are suppressed.
Configure the maximum and minimum intervals for sending RA messages	ipv6 nd ra interval <i>max-interval-value</i> <i>min-interval-value</i>	Optional By default, the maximum interval for sending RA messages is 600 seconds, and the minimum interval is 200 seconds. The device sends RA messages at intervals of a random value between the maximum interval and the minimum interval. The minimum interval should be less than or equal to 0.75 times the maximum interval.
Configure the prefix information options in RA messages	ipv6 nd ra prefix { <i>ipv6-address prefix-length</i> <i>ipv6-address/prefix-length</i> } <i>valid-lifetime</i> <i>preferred-lifetime</i> [no-autoconfig [off-link]*	Optional By default, no prefix information is configured in RA messages and the IPv6 address of the interface sending RA messages is used as the prefix information.
Set the M flag bit to 1	ipv6 nd autoconfig managed-address-flag	Optional By default, the M flag bit is set to 0, that is, hosts acquire IPv6 addresses through stateless autoconfiguration.
Set the O flag bit to 1.	ipv6 nd autoconfig other-flag	Optional By default, the O flag bit is set to 0, that is, hosts acquire other information through stateless autoconfiguration.
Configure the router lifetime in RA messages	ipv6 nd ra router-lifetime <i>value</i>	Optional 1,800 seconds by default.
Set the retrans timer	ipv6 nd ns retrans-timer <i>value</i>	Optional By default, the local interface sends NS messages at intervals of 1,000 milliseconds and the Retrans Timer field in RA messages sent by the local interface is equal to 0.

To do...	Use the command...	Remarks
Set the reachable time	ipv6 nd nud reachable-time <i>value</i>	Optional By default, the neighbor reachable time on the local interface is 30,000 milliseconds and the Reachable Timer field in RA messages is 0.



CAUTION: The maximum interval for sending RA messages should be less than or equal to the router lifetime in RA messages.

Configuring the Number of Attempts to Send an NS Message for DAD

An interface sends a neighbor solicitation (NS) message for DAD after acquiring an IPv6 address. If the interface does not receive a response within a specified time (determined by the **ipv6 nd ns retrans-timer** command), it continues to send an NS message. If it still does not receive a response after the number of attempts to send an NS message reaches the maximum, the acquired address is considered available.

Follow these steps to configure the attempts to send an NS message for DAD:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure the number of attempts to send an NS message for DAD	ipv6 nd dad attempts <i>value</i>	Optional 1 by default. When the <i>value</i> argument is set to 0, DAD is disabled.

Configuring PMTU Discovery

Configuring a Static PMTU for a Specified IPv6 Address

You can configure a static PMTU for a specified destination IPv6 address. When a source host sends packets through an interface, it compares the interface MTU with the static PMTU of the specified destination IPv6 address. If the packet size is larger than the smaller one between the two values, the host fragments the packet according to the smaller value.

Follow these steps to configure a static PMTU for a specified address:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure a static PMTU for a specified IPv6 address	ipv6 pathmtu <i>ipv6-address</i> [<i>value</i>]	Required By default, no static PMTU is configured.

Configuring the Aging Time for PMTU

After the MTU of the path from the source host to the destination host is dynamically determined (refer to "IPv6 PMTU Discovery" on page 507), the source

host sends subsequent packets to the destination host on basis of this MTU. After the aging time expires, the dynamically determined PMTU is removed and the source host re-determines an MTU to send packets through the PMTU mechanism.

The aging time is invalid for static PMTU.

Follow these steps to configure the aging time for PMTU:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure aging time for PMTU	ipv6 pathmtu age <i>age-time</i>	Optional 10 minutes by default.

Configuring IPv6 TCP Properties

The IPv6 TCP properties you can configure include:

- **synwait timer:** When a SYN packet is sent, the synwait timer is triggered. If no response packet is received before the synwait timer expires, the IPv6 TCP connection establishment fails.
- **finwait timer:** When the IPv6 TCP connection status is FIN_WAIT_2, the finwait timer is triggered. If no packet is received before the finwait timer expires, the IPv6 TCP connection is terminated. If a FIN packet is received, the IPv6 TCP connection status becomes TIME_WAIT. If other packets are received, the finwait timer is reset from the last received packet and the connection is terminated after the finwait timer expires.
- **Size of the IPv6 TCP sending/receiving buffer.**

Follow these steps to configure IPv6 TCP properties:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Set the finwait timer of IPv6 TCP packets	tcp ipv6 timer fin-timeout <i>wait-time</i>	Optional 675 seconds by default.
Set the synwait timer of IPv6 TCP packets	tcp ipv6 timer syn-timeout <i>wait-time</i>	Optional 75 seconds by default.
Set the size of the IPv6 TCP sending/receiving buffer	tcp ipv6 window size	Optional 8 KB by default.

Configuring ICMPv6 Packet Sending

Configuring the Maximum ICMPv6 Error Packets Sent in an Interval

If too many ICMPv6 error packets are sent within a short time in a network, network congestion may occur. To avoid network congestion, you can control the maximum number of ICMPv6 error packets sent within a specified time, currently by adopting the token bucket algorithm.

You can set the capacity of a token bucket, namely, the number of tokens in the bucket. In addition, you can set the update period of the token bucket, namely, the interval for updating the number of tokens in the token bucket to the

configured capacity. One token allows one ICMPv6 error packet to be sent. Each time an ICMPv6 error packet is sent, the number of tokens in a token bucket decreases by 1. If the number of ICMPv6 error packets successively sent exceeds the capacity of the token bucket, subsequent ICMPv6 error packets cannot be sent out until the number of tokens in the token bucket is updated and new tokens are added to the bucket.

Follow these steps to configure the capacity and update period of the token bucket:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the capacity and update period of the token bucket	ipv6 icmp-error { bucket bucket-size ratelimit interval } *	Optional By default, the capacity of a token bucket is 10 and the update period is 100 milliseconds. That is, at most 10 IPv6 ICMP error packets can be sent within these 100 milliseconds. The update period "0" indicates that the number of ICMPv6 error packets sent is not restricted.

Enable Sending of Multicast Echo Replies

If hosts are capable of relying multicast echo requests, Host A can attack Host B by sending an echo request with the source being Host B to a multicast address, then all the hosts in the multicast group will send echo replies to Host B. Therefore, a device is disabled from replying multicast echo requests by default.

Follow these steps to enable sending of multicast echo replies:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable sending of multicast echo replies	ipv6 icmpv6 multicast-echo-reply enable	Not enabled by default.

Configuring IPv6 DNS

Configuring Static IPv6 Domain Name Resolution

Configuring static IPv6 domain name resolution is to establish the mapping between host name and IPv6 address. When applying such applications as Telnet, you can directly use a host name and the system will resolve the host name into an IPv6 address. Each host name can correspond to only one IPv6 address.

Follow these steps to configure static IPv6 domain name resolution:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure a host name and the corresponding IPv6 address	ipv6 host <i>hostname</i> <i>ipv6-address</i>	Required

Configuring Dynamic IPv6 Domain Name Resolution

If you want to use the dynamic domain name function, you can use the following command to enable the dynamic domain name resolution function. In addition, you should configure a DNS server so that a query request message can be sent to the correct server for resolution. The system can support at most six DNS servers.

You can configure a DNS suffix so that you only need to enter some fields of a domain name and the system can automatically add the preset suffix for address resolution. The system can support at most 10 DNS suffixes.

Follow these steps to configure dynamic IPv6 domain name resolution:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the dynamic domain name resolution function	dns resolve	Required Disabled by default.
Configure an IPv6 DNS server	dns server ipv6 <i>ipv6-address</i> [<i>interface-type</i> <i>interface-number</i>]	Required If the IPv6 address of the DNS server is a link-local address, you need to specify a value for <i>interface-type</i> and <i>interface-number</i> .
Configure the DNS suffix.	dns domain <i>domain-name</i>	Required By default, no DN suffix is configured, that is, the domain name is resolved according to the input information.



The **dns resolve** and **dns domain** commands are the same as those of IPv4 DNS. For details about the commands, refer to "DNS Configuration" on page 971.

Displaying and Maintaining IPv6 Basics Configuration

To do...	Use the command...	Remarks
Display DNS suffix information	display dns domain [dynamic]	Available in any view
Display IPv6 dynamic domain name cache information.	display dns ipv6 dynamic-host	
Display IPv6 DNS server information	display dns ipv6 server [dynamic]	
Display the IPv6 FIB entries	display ipv6 fib [<i>ipv6-address</i>]	
Display the mappings between host names and IPv6 addresses in the static DNS database.	display ipv6 host	
Display the IPv6 information of an interface	display ipv6 interface [brief] [<i>interface-type</i> [<i>interface-number</i>]]	
Display neighbor information	display ipv6 neighbors { <i>ipv6-address</i> all dynamic interface <i>interface-type</i> <i>interface-number</i> static vlan <i>vlan-id</i> } [[begin exclude include] <i>string</i>]	

To do...	Use the command...	Remarks
Display the total number of neighbor entries satisfying the specified conditions	display ipv6 neighbors { all dynamic interface <i>interface-type interface-number</i> static vlan <i>vlan-id</i> } count	Available in any view
Display the PMTU information of an IPv6 address	display ipv6 pathmtu { <i>ipv6-address</i> all dynamic static }	
Display information related to a specified socket	display ipv6 socket [sockettype <i>socket-type</i>] [<i>task-id socket-id</i>]	
Display the statistics of IPv6 packets and ICMPv6 packets	display ipv6 statistics	
Display the IPv6 TCP connection statistics	display tcp ipv6 statistics	
Display the IPv6 TCP connection status	display tcp ipv6 status	
Display the IPv6 UDP connection statistics	display udp ipv6 statistics	
Clear IPv6 dynamic domain name cache information	reset dns ipv6 dynamic-host	Available in user view
Clear IPv6 neighbor information	reset ipv6 neighbors { all dynamic interface <i>interface-type interface-number</i> static }	
Clear the corresponding PMTU	reset ipv6 pathmtu { all static dynamic }	
Clear the statistics of IPv6 and ICMPv6 packets	reset ipv6 statistics	
Clear all IPv6 TCP connection statistics	reset tcp ipv6 statistics	
Clear the statistics of all IPv6 UDP packets	reset udp ipv6 statistics	



The **display dns domain** command is the same as the one of IPv4 DNS. For details about the commands, refer to “DNS Configuration” on page 971.

IPv6 Configuration Example

Network requirements

Two switches are directly connected through two Ethernet ports. The Ethernet ports belong to VLAN 2. Configure different types of IPv6 addresses for VLAN-interface 2 respectively on Switch A and Switch B to verify the connectivity between two switches. The IPv6 prefix in the EUI-64 format is 2001::/64. Specify the aggregatable global unicast address of Switch A as 3001::1/64, and the aggregatable global unicast address of Switch B as 3001::2/64.

Network diagram

Figure 155 Network diagram for IPv6 address configuration



Configuration procedure

- Configuration on Switch A

Enable the IPv6 packet forwarding function.

```
<SwitchA> system-view
[SwitchA] ipv6
```

Configure VLAN-interface 2 to automatically generate a link-local address.

```
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ipv6 address auto link-local
```

Configure an EUI-64 address for VLAN-interface 2.

```
[SwitchA-Vlan-interface2] ipv6 address 2001::/64 eui-64
```

Specify an aggregatable global unicast address for VLAN-interface 2.

```
[SwitchA-Vlan-interface2] ipv6 address 3001::1/64
```

Allow VLAN-interface 2 to advertise RA messages.

```
[SwitchA-Vlan-interface2] undo ipv6 nd ra halt
```

■ Configuration on Switch B

Enable the IPv6 packet forwarding function.

```
<SwitchB> system-view
[SwitchB] ipv6
```

Configure VLAN-interface 2 to automatically generate a link-local address.

```
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ipv6 address auto link-local
```

Configure an EUI-64 address for VLAN-interface 2.

```
[SwitchB-Vlan-interface2] ipv6 address 2001::/64 eui-64
```

Configure an aggregatable global unicast address for VLAN-interface 2.

```
[SwitchB-Vlan-interface2] ipv6 address 3001::2/64
```

Verification

Display the IPv6 information of the interface on Switch A.

```
[SwitchA-Vlan-interface2] display ipv6 interface vlan-interface 2
Vlan-interface2 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::20F:E2FF:FE49:8048
Global unicast address(es):
  2001::20F:E2FF:FE49:8048, subnet is 2001::/64
  3001::1, subnet is 3001::/64
Joined group address(es):
  FF02::1:FF00:1
  FF02::1:FF49:8048
  FF02::2
  FF02::1
MTU is 1500 bytes
```

```

ND DAD is enabled, number of DAD attempts: 1
ND reachable time is 30000 milliseconds
ND retransmit interval is 1000 milliseconds
Hosts use stateless autoconfig for addresses

```

Display the IPv6 information of the interface on Switch B.

```

[SwitchB-Vlan-interface2] display ipv6 interface vlan-interface 2
Vlan-interface2 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::20F:E2FF:FE00:1
Global unicast address(es):
  2001::20F:E2FF:FE00:1, subnet is 2001::/64
  3001::2, subnet is 3001::/64
Joined group address(es):
  FF02::1:FF00:2
  FF02::1:FF00:1
  FF02::2
  FF02::1
MTU is 1500 bytes
ND DAD is enabled, number of DAD attempts: 1
ND reachable time is 30000 milliseconds
ND retransmit interval is 1000 milliseconds
Hosts use stateless autoconfig for addresses

```

From Switch A, ping the link-local address, EUI-64 address, and aggregatable global unicast address respectively. If the configurations are correct, the three types of IPv6 addresses above can be pinged.



CAUTION: When you ping a link-local address, you should use the “-i” parameter to specify an interface for the link-local address.

```

[SwitchA-Vlan-interface2] ping ipv6 FE80::20F:E2FF:FE00:1 -i vlan-interface2
PING FE80::20F:E2FF:FE00:1 : 56 data bytes, press CTRL_C to break
  Reply from FE80::20F:E2FF:FE00:1
  bytes=56 Sequence=1 hop limit=255 time = 80 ms
  Reply from FE80::20F:E2FF:FE00:1
  bytes=56 Sequence=2 hop limit=255 time = 60 ms
  Reply from FE80::20F:E2FF:FE00:1
  bytes=56 Sequence=3 hop limit=255 time = 60 ms
  Reply from FE80::20F:E2FF:FE00:1
  bytes=56 Sequence=4 hop limit=255 time = 70 ms
  Reply from FE80::20F:E2FF:FE00:1
  bytes=56 Sequence=5 hop limit=255 time = 60 ms

--- FE80::20F:E2FF:FE00:1 ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 60/66/80 ms
[SwitchA-Vlan-interface2] ping ipv6 2001::20F:E2FF:FE00:1
PING 2001::20F:E2FF:FE00:1 : 56 data bytes, press CTRL_C to break
  Reply from 2001::20F:E2FF:FE00:1
  bytes=56 Sequence=1 hop limit=255 time = 40 ms
  Reply from 2001::20F:E2FF:FE00:1
  bytes=56 Sequence=2 hop limit=255 time = 70 ms
  Reply from 2001::20F:E2FF:FE00:1
  bytes=56 Sequence=3 hop limit=255 time = 60 ms
  Reply from 2001::20F:E2FF:FE00:1
  bytes=56 Sequence=4 hop limit=255 time = 60 ms
  Reply from 2001::20F:E2FF:FE00:1
  bytes=56 Sequence=5 hop limit=255 time = 60 ms

```

```

--- 2001::20F:E2FF:FE00:1 ping statistics ---
 5 packet(s) transmitted
 5 packet(s) received
 0.00% packet loss
 round-trip min/avg/max = 40/58/70 ms

[SwitchA-Vlan-interface2] ping ipv6 3001::2
PING 3001::2 : 56 data bytes, press CTRL_C to break
Reply from 3001::2
 bytes=56 Sequence=1 hop limit=255 time = 50 ms
Reply from 3001::2
 bytes=56 Sequence=2 hop limit=255 time = 60 ms
Reply from 3001::2
 bytes=56 Sequence=3 hop limit=255 time = 60 ms
Reply from 3001::2
 bytes=56 Sequence=4 hop limit=255 time = 70 ms
Reply from 3001::2
 bytes=56 Sequence=5 hop limit=255 time = 60 ms

--- 3001::2 ping statistics ---
 5 packet(s) transmitted
 5 packet(s) received
 0.00% packet loss
 round-trip min/avg/max = 50/60/70 ms

```

Troubleshooting IPv6 Basics Configuration

Symptom

The peer IPv6 address cannot be pinged.

Solution

- Use the **display current-configuration** command in any view or the **display this** command in system view to check that the IPv6 packet forwarding function is enabled.
- Use the **display ipv6 interface** command in any view to check that the IPv6 address of the interface is correct and that the interface is up.
- Use the **debugging ipv6 packet** command in user view to enable the debugging for IPv6 packets and make judgment according to the debugging information.

39

DUAL STACK CONFIGURATION

When configuring dual stack, go to these sections for information you are interested in:

- "Dual Stack Overview" on page 521
- "Configuring Dual Stack" on page 521

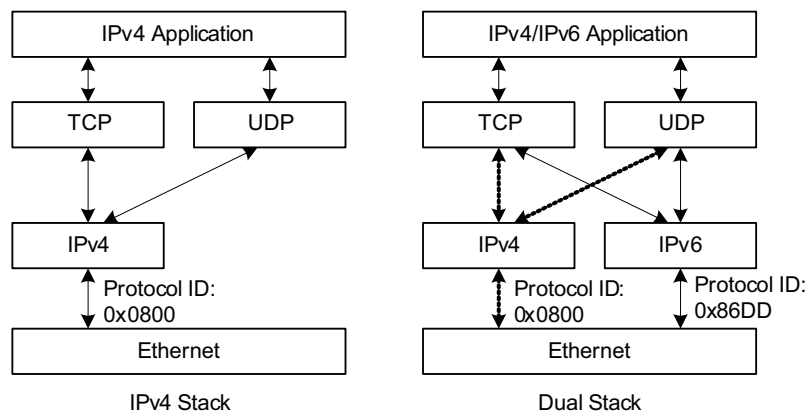
Dual Stack Overview

Dual stack is the most direct approach to making IPv6 nodes compatible with IPv4 nodes. The best way for an IPv6 node to be compatible with an IPv4 node is to maintain a complete IPv4 stack. A network node that supports both IPv4 and IPv6 is called a dual stack node. A dual stack node configured with an IPv4 address and an IPv6 address can have both IPv4 and IPv6 packets transmitted.

For an upper layer application supporting both IPv4 and IPv6, either TCP or UDP can be selected at the transport layer, while IPv6 stack is preferred at the network layer.

Figure 156 illustrates the IPv4/IPv6 dual stack in relation to the IPv4 stack.

Figure 156 IPv4/IPv6 dual stack in relation to IPv4 stack (on Ethernet)



Configuring Dual Stack

You must enable the IPv6 packet forwarding function before dual stack. Otherwise, the device cannot forward IPv6 packets even if IPv6 addresses are configured for interfaces.

Follow these steps to configure dual stack on a gateway:

To do...	Use the command...	Remarks
Enter system view	<code>system-view</code>	-

To do...			Use the command...	Remarks
Enable the IPv6 packet forwarding function			ipv6	Required Disabled by default.
Enter interface view			interface <i>interface-type</i> <i>interface-number</i>	-
Configure an IPv4 address for the interface			ip address <i>ip-address</i> { <i>mask</i> <i>mask-length</i> } [sub]	Required By default, no IP address is configured.
Configure an IPv6 address on the interface	Configure IPv6 global unicast address or local address	Manually specify an IPv6 address	ipv6 address { <i>ipv6-address</i> <i>prefix-length</i> <i>ipv6-address/prefix-length</i> }	Use either command. By default, no local address or global unicast address is configured on an interface
		Configure an IPv6 address in the EUI-64 format	ipv6 address <i>ipv6-address/prefix-length</i> eui-64	
Configure IPv6 link-local address	Configure IPv6 link-local address	Automatically create an IPv6 link-local address	ipv6 address auto link-local	Optional By default, after you configured an IPv6 local address or global unicast address, a link local address is automatically created.
		Manually specify an IPv6 link-local address	ipv6 address <i>ipv6-address</i> link-local	

40

TUNNELING CONFIGURATION

When configuring tunneling, go to these sections for information you are interested in:

- "Introduction to Tunneling" on page 523
- "Tunneling Configuration Task List" on page 526
- "Configuring IPv6 Manual Tunnel" on page 526
- "Configuring 6to4 Tunnel" on page 530
- "Configuring ISATAP Tunnel" on page 535
- "Displaying and Maintaining Tunneling Configuration" on page 538
- "Troubleshooting Tunneling Configuration" on page 538

Introduction to Tunneling

Tunneling is an encapsulation technology, which utilizes one network transport protocol to encapsulate packets of another network transport protocol and transfer them over the network. A tunnel is a virtual point-to-point connection. In practice, the virtual interface that supports only point-to-point connections is called tunnel interface. One tunnel provides one channel to transfer encapsulated packets. Packets can be encapsulated and decapsulated at both ends of a tunnel. Tunneling refers to the whole process from data encapsulation to data transfer to data decapsulation.



NTP-related commands are available in tunnel interface view on 3Com Switch 4800G Family, but NTP features cannot be enabled after you execute the NTP commands. For related information about NTP, refer to "NTP Configuration" on page 947.

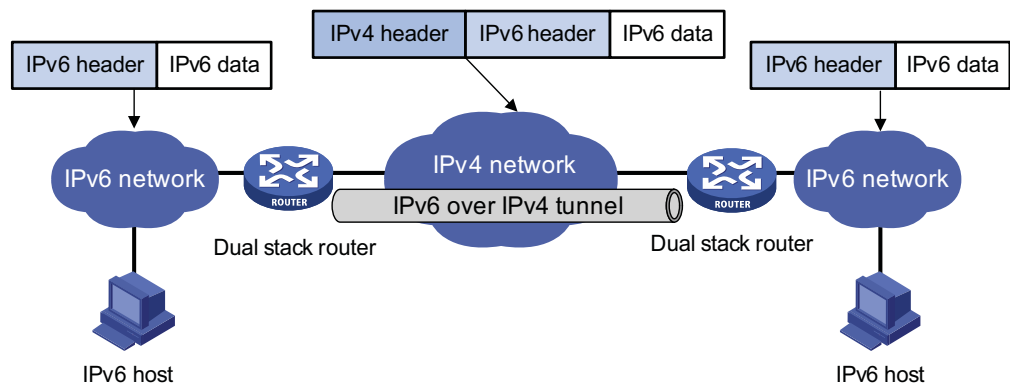
IPv6 over IPv4 Tunnel

Principle

The IPv6 over IPv4 tunneling mechanism encapsulates an IPv4 header in IPv6 data packets so that IPv6 packets can pass an IPv4 network through a tunnel to realize interworking between isolated IPv6 networks, as shown in Figure 157.



CAUTION: *The devices at both ends of an IPv6 over IPv4 tunnel must support IPv4/IPv6 dual stack.*

Figure 157 Principle of IPv6 over IPv4 tunnel

The IPv6 over IPv4 tunnel processes packets in the following way:

- 1 A host in the IPv6 network sends an IPv6 packet to the device at the source end of the tunnel.
- 2 After determining according to the routing table that the packet needs to be forwarded through the tunnel, the device at the source end of the tunnel encapsulates the IPv6 packet with an IPv4 header and forwards it through the physical interface of the tunnel.
- 3 The encapsulated packet goes through the tunnel to reach the device at the destination end of the tunnel. The device at the destination end decapsulates the packet if the destination address of the encapsulated packet is the device itself.
- 4 The destination device forwards the packet according to the destination address in the decapsulated IPv6 packet. If the destination address is the device itself, the device forwards the IPv6 packet to the upper-layer protocol for processing.

Configured tunnel and automatic tunnel

An IPv6 over IPv4 tunnel can be established between hosts, between hosts and devices, and between devices. The tunnel destination needs to forward packets if the tunnel destination is not the eventual destination of the IPv6 packet.

According to the way the IPv4 address of the tunnel destination is acquired, tunnels are divided into configured tunnel and automatic tunnel.

- If the IPv4 address of the tunnel destination cannot be acquired from the destination address of the IPv6 packet, it needs to be configured manually. Such a tunnel is called a configured tunnel.
- If the IPv4 address is embedded into the IPv6 address, the IPv4 address of the tunnel destination can automatically be acquired from the destination address of the IPv6 packet. Such a tunnel is called an automatic tunnel.

Type

According to the way an IPv6 packet is encapsulated, IPv6 over IPv4 tunnels are divided into the following types:

- IPv6 manual tunnel
- 6to4 tunnel
- ISATAP tunnel

Among the above tunnels, the IPv6 manual tunnel is a configured tunnel, while the 6to4 tunnel, and intra-site automatic tunnel address protocol (ISATAP) tunnel are automatic tunnels.

1 IPv6 manually configured tunnel

A manually configured tunnel is a point-to-point link. One link is a separate tunnel. The IPv6 manually configured tunnels provide stable connections requiring regular secure communication between two border routers or between a border router and a host for access to remote IPv6 networks.

2 6to4 tunnel

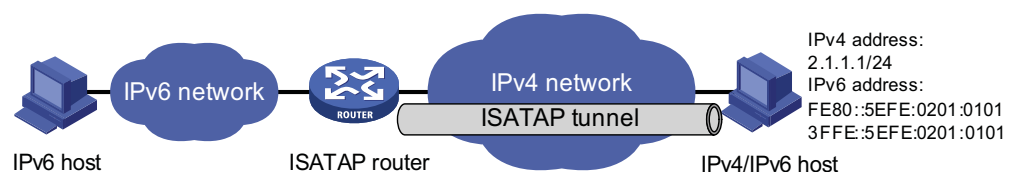
An automatic 6to4 tunnel is a point-to-multipoint tunnel and is used to connect multiple isolated IPv6 networks over an IPv4 network to remote IPv6 networks. The embedded IPv4 address in an IPv6 address is used to automatically acquire the destination of the tunnel. The automatic 6to4 tunnel adopts 6to4 addresses. The address format is 2002:abcd:efgh:subnet number::interface ID/64, where abcd:efgh represents the 32-bit source IPv4 address of the 6to4 tunnel, in hexadecimal notation. For example, 1.1.1.1 can be represented by 0101:0101. The tunnel destination is automatically determined by the embedded IPv4 address, which makes it easy to create a 6to4 tunnel.

Since the 16-bit subnet number of the 64-bit address prefix in 6to4 addresses can be customized and the first 48 bits in the address prefix are fixed by a permanent value and the IPv4 address of the tunnel source or destination, it is possible that IPv6 packets can be forwarded by the tunnel.

3 ISATAP tunnel

With the application of the IPv6 technology, there will be more and more IPv6 hosts in the existing IPv4 network. The ISATAP tunneling technology provides a satisfactory solution for IPv6 application. An ISATAP tunnel is a point-to-point automatic tunnel. The destination of a tunnel can automatically be acquired from the embedded IPv4 address in the destination address of an IPv6 packet. When an ISATAP tunnel is used, the destination address of an IPv6 packet and the IPv6 address of a tunnel interface both adopt special addresses: ISATAP addresses. The ISATAP address format is prefix(64bit):0:5EFE:ip-address. The ip-address is in the form of a.b.c.d or abcd:efgh, where abcd:efgh represents a 32-bit source IPv4 address. Through the embedded IPv4 address, an ISATAP tunnel can automatically be created to transfer IPv6 packets. The ISATAP tunnel is mainly used for connection between IPv6 routers or between a host and an IPv6 router over an IPv4 network.

Figure 158 Principle of ISATAP tunnel



Tunneling Configuration Task List

Complete the following tasks to configure the tunneling feature:

Task	Remarks
Configuring IPv6 over IPv4 GRE tunnel	“Configuring IPv6 Manual Tunnel” on page 526 Optional
	“Configuring 6to4 Tunnel” on page 530 Optional
	“Configuring ISATAP Tunnel” on page 535 Optional

Configuring IPv6 Manual Tunnel

Configuration Prerequisites

IP addresses are configured for interfaces such as the VLAN interface and loopback interface on the device. These interfaces serve as the source interfaces of tunnel interfaces to ensure that the tunnel destination addresses are reachable.

Configuration Procedure

Follow these steps to configure an IPv6 manual tunnel:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable IPv6	ipv6	Required By default, the IPv6 packet forwarding function is disabled.
Create a tunnel interface and enter tunnel interface view	interface tunnel <i>number</i>	Required By default, there is no tunnel interface on the device.
Configure an IPv6 address for the tunnel interface	ipv6 address { <i>ipv6-address prefix-length</i> <i>ipv6-address/prefix-length</i> } ipv6 address <i>ipv6-address/prefix-length eui-64</i>	Required Use any command. By default, no IPv6 global unicast address or site-local address is configured for the tunnel interface.
	Configure a link-local IPv6 address ipv6 address auto link-local ipv6 address <i>ipv6-address link-local</i>	Optional A link-local address will automatically be created when an IPv6 global unicast address or site-local address is configured.
Specify the IPv6 manual tunnel mode	tunnel-protocol ipv6-ipv4	Required By default, the tunnel mode is manual. The same tunnel type should be configured at both ends of the tunnel. Otherwise, packet delivery will fail.

To do...	Use the command...	Remarks
Configure a source address or interface for the tunnel	source { <i>ip-address</i> <i>interface-type</i> <i>interface-number</i> }	Required By default, no source address or interface is configured for the tunnel.
Configure a destination address for the tunnel	destination <i>ip-address</i>	Required By default, no destination address is configured for the tunnel.
Reference an aggregation group	aggregation-group <i>aggregation-group-id</i>	Required By default, no link aggregation group ID is referenced.

**CAUTION:**

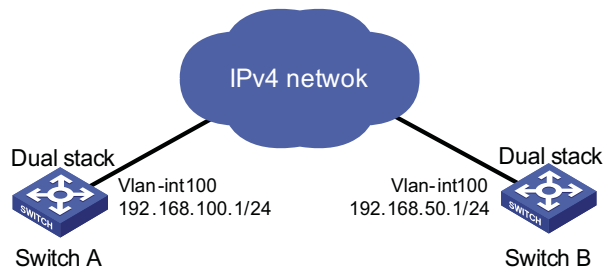
- *After a tunnel interface is deleted, all the above features configured on the tunnel interface will be deleted.*
- *If the addresses of the tunnel interfaces at the two ends of a tunnel are not in the same network segment, a forwarding route through the tunnel to the peer must be configured so that the encapsulated packet can be forwarded normally. The route can be a static or dynamic route. IP addresses must be configured at both ends of the tunnel. For detailed configuration, refer to "Routing Policy Configuration" on page 415 or "Routing Policy Configuration" on page 489.*
- *When you configure a static route, you need to configure a route to the destination address (the destination IPv6 address of the packet, instead of the IPv4 address of the tunnel destination) and set the next-hop to the tunnel interface number or network address at the local end of the tunnel. Such configurations must be performed at both ends of the tunnel.*
- *Before configuring dynamic routes, you must enable the dynamic routing protocol on the tunnel interfaces at both ends. For related configurations, refer to "Routing Policy Configuration" on page 489.*
- *Before referencing a link aggregation group on the tunnel interface to receive and send packets, make sure that the aggregation group has been configured. Otherwise, the tunnel interface will not be up to communicate.*

Configuration Example Network requirements

Two IPv6 networks are connected through an IPv6 manual tunnel between Switch A and Switch B. As shown in Figure 159, VLAN-interface 100 on Switch A can communicate with VLAN-interface 100 on Switch B normally via an IPv4 route.

Network diagram

Figure 159 Network diagram for an IPv6 manual tunnel



Configuration procedure

■ Configuration on Switch A

Enable IPv6.

```
<SwitchA> system-view
[SwitchA] ipv6
```

Configure a link aggregation group. Disable STP on the port before adding it into the link aggregation group.

```
[SwitchA] link-aggregation group 1 mode manual
[SwitchA] link-aggregation group 1 service-type tunnel
[SwitchA] interface GigabitEthernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] stp disable
[SwitchA-GigabitEthernet1/0/1] port link-aggregation group 1
[SwitchA-GigabitEthernet1/0/1] quit
```

Configure an IPv4 address for VLAN-interface 100.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port GigabitEthernet 1/0/2
[SwitchA-vlan100] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] ip address 192.168.100.1 255.255.255.0
[SwitchA-Vlan-interface100] quit
```

Configure a manual IPv6 tunnel.

```
[SwitchA] interface tunnel 0
[SwitchA-Tunnel0] ipv6 address 3001::1/64
[SwitchA-Tunnel0] source vlan-interface 100
[SwitchA-Tunnel0] destination 192.168.50.1
[SwitchA-Tunnel0] tunnel-protocol ipv6-ipv4
```

Configure the tunnel to reference link aggregation group 1 in tunnel interface view.

```
[SwitchA-Tunnel0] aggregation-group 1
```

■ Configuration on Switch B

Enable IPv6.

```
<SwitchB> system-view
[SwitchB] ipv6
```

Configure a link aggregation group. Disable STP on the port before adding it into the link aggregation group.

```
[SwitchB] link-aggregation group 1 mode manual
[SwitchB] link-aggregation group 1 service-type tunnel
[SwitchB] interface GigabitEthernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] stp disable
[SwitchB-GigabitEthernet1/0/1] port link-aggregation group 1
[SwitchB-GigabitEthernet1/0/1] quit
```

Configure an IPv4 address for VLAN-interface 100.

```
[SwitchB] vlan 100
[SwitchB-vlan100] port GigabitEthernet 1/0/2
[SwitchB-vlan100] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ip address 192.168.50.1 255.255.255.0
[SwitchB-Vlan-interface100] quit
```

Configure an IPv6 manual tunnel.

```
[SwitchB] interface tunnel 0
[SwitchB-Tunnel0] ipv6 address 3001::2/64
[SwitchB-Tunnel0] source vlan-interface 100
[SwitchB-Tunnel0] destination 192.168.100.1
[SwitchB-Tunnel0] tunnel-protocol ipv6-ipv4
```

Configure the tunnel to reference link aggregation group 1 in tunnel interface view.

```
[SwitchB-Tunnel0] aggregation-group 1
```

Configuration verification

After the above configurations, display the status of the tunnel interfaces on Switch A and Switch B, respectively.

```
[SwitchA] display ipv6 interface tunnel 0
Tunnel0 current state :UP
Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::C0A8:6401
Global unicast address(es):
  3001::1, subnet is 3001::/64
Joined group address(es):
  FF02::1:FFA8:6401
  FF02::1:FF00:1
  FF02::2
  FF02::1
MTU is 1500 bytes
ND reachable time is 30000 milliseconds
ND retransmit interval is 1000 milliseconds
Hosts use stateless autoconfig for addresses
```

```
[SwitchB] display ipv6 interface tunnel 0
Tunnel0 current state :UP
```

```

Line protocol current state :UP
IPv6 is enabled, link-local address is FE80::C0A8:3201
Global unicast address(es):
  3001::2, subnet is 3001::/64
Joined group address(es):
  FF02::1:FFA8:3201
  FF02::1:FF00:2
  FF02::2
  FF02::1
MTU is 1500 bytes
ND reachable time is 30000 milliseconds
ND retransmit interval is 1000 milliseconds
Hosts use stateless autoconfig for addresses

```

Ping the IPv6 address of the peer tunnel interface from Switch A.

```

[SwitchA] ping ipv6 3001::2
PING 3001::2 : 56 data bytes, press CTRL_C to break
  Reply from 3001::2
    bytes=56 Sequence=1 hop limit=64 time = 31 ms
  Reply from 3001::2
    bytes=56 Sequence=2 hop limit=64 time = 16 ms
  Reply from 3001::2
    bytes=56 Sequence=3 hop limit=64 time = 1 ms
  Reply from 3001::2
    bytes=56 Sequence=4 hop limit=64 time = 15 ms
  Reply from 3001::2
    bytes=56 Sequence=5 hop limit=64 time = 15 ms

--- 3001::2 ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 1/15/31 ms

```

Configuring 6to4 Tunnel

Configuration Prerequisites

IP addresses are configured for interfaces such as VLAN interface and loopback interface on the device. Such an interface can serve as the source interface of the tunnel to ensure that the tunnel destination address is reachable.

Configuration Procedure

Follow these steps to configure a 6to4 tunnel:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable IPv6	ipv6	Required By default, the IPv6 packet forwarding function is disabled.

To do...	Use the command...	Remarks
Create a tunnel interface and enter tunnel interface view	interface tunnel <i>number</i>	Required By default, there is no tunnel interface on the device.
Configure an IPv6 address for the tunnel interface	Configure an IPv6 global unicast address or site-local address ipv6 address { <i>ipv6-address prefix-length</i> <i>ipv6-address/prefix-length</i> }	Required. Use either command. By default, no IPv6 global unicast address or site-local address is configured for the tunnel interface.
	Configure an IPv6 link-local address ipv6 address auto link-local ipv6 address <i>ipv6-address link-local</i>	Optional By default, a link-local address will automatically be generated when an IPv6 global unicast address or site-local address is configured.
Set a 6to4 tunnel	tunnel-protocol ipv6-ipv4 6to4	Required By default, the tunnel mode is manual. The same tunnel type should be configured at both ends of the tunnel. Otherwise, packet delivery will fail.
Configure a source address or interface for the tunnel	source { <i>ip-address</i> <i>interface-type interface-number</i> }	Required By default, no source address or interface is configured for the tunnel.
Reference a link aggregation group	aggregation-group <i>aggregation-group-id</i>	Required By default, no link aggregation group ID is referenced.

**CAUTION:**

- *Only one automatic tunnel can be configured at the same tunnel source.*
- *No destination address needs to be configured for an automatic tunnel because the destination address can automatically be obtained from the IPv4 address embedded in the IPv4-compatible IPv6 address.*
- *When you create a tunnel interface on a device, the slot of the tunnel interface should be that of the source interface, namely, the interface sending packets. In this way, the forwarding efficiency can be improved.*
- *If the addresses of the tunnel interfaces at the two ends of a tunnel are not in the same network segment, a forwarding route through the tunnel to the peer must be configured so that the encapsulated packet can be forwarded normally. You can configure static or dynamic routes. You should perform this configuration at both ends of the tunnel.*
- *The automatic tunnel interfaces encapsulated with the same protocol cannot share the same source IP address.*
- *Automatic tunnels do not support dynamic routing.*

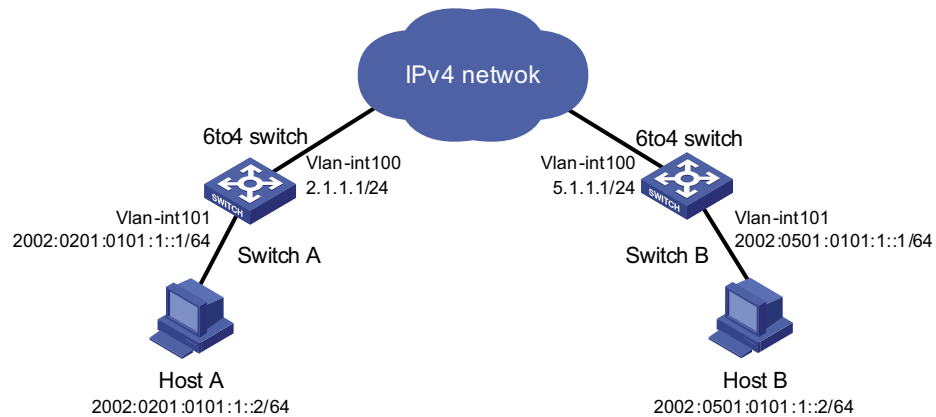
- When you configure a static route, you need to configure a route to the destination address (the destination IP address of the packet, instead of the IPv4 address of the tunnel destination) and set the next-hop to the tunnel interface number or network address at the local end of the tunnel. Such a route must be configured at both ends of the tunnel.
- Before referencing a link aggregation group on the tunnel interface to receive and send packets, make sure that the aggregation group has been configured. Otherwise, the tunnel interface will not be up to communicate.

Configuration Example Network requirements

Isolated IPv6 networks are interconnected through a 6to4 tunnel over the IPv4 network.

Network diagram

Figure 160 Network diagram for a 6to4 tunnel



Configuration procedure

- Configuration on Switch A

Enable IPv6.

```
<SwitchA> system-view
[SwitchA] ipv6
```

Configure a link aggregation group. Disable STP on the port before adding it into the link aggregation group.

```
[SwitchA] link-aggregation group 1 mode manual
[SwitchA] link-aggregation group 1 service-type tunnel
[SwitchA] interface GigabitEthernet 1/0/1
[SwitchA-GigabitEthernet1/0/1] stp disable
[SwitchA-GigabitEthernet1/0/1] port link-aggregation group 1
[SwitchA-GigabitEthernet1/0/1] quit
```

Configure an IPv4 address for VLAN-interface 100.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port GigabitEthernet 1/0/2
[SwitchA-vlan100] quit
[SwitchA] interface vlan-interface 100
```



```
[SwitchA-Vlan-interface100] ip address 2.1.1.1 24
[SwitchA-Vlan-interface100] quit
```

Configure a route to VLAN-interface 100 of Switch B. (Here the next-hop address of the static route is represented by [nexthop]. In practice, you should configure the real next-hop address according to the network.)

```
[SwitchA] ip route-static 5.1.1.1 24 [nexthop]
```

Configure an IPv6 address for VLAN-interface 101.

```
[SwitchA] vlan 101
[SwitchA-vlan101] port GigabitEthernet 1/0/3
[SwitchA-vlan101] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] ipv6 address 2002:0201:0101:1::1/64
[SwitchA-Vlan-interface101] quit
```

Configure a 6to4 tunnel.

```
[SwitchA] interface tunnel 0
[SwitchA-Tunnel0] ipv6 address 2002:201:101::1/64
[SwitchA-Tunnel0] source vlan-interface 100
[SwitchA-Tunnel0] tunnel-protocol ipv6-ipv4 6to4
[SwitchA-Tunnel0] quit
```

Configure the tunnel to reference link aggregation group 1 in tunnel interface view.

```
[SwitchA-Tunnel0] aggregation-group 1
[SwitchA-Tunnel0] quit
```

Configure a static route whose destination address is 2002::/16 and next-hop is the tunnel interface.

```
[SwitchA] ipv6 route-static 2002:: 16 tunnel 0
```

■ Configuration on Switch B

Enable IPv6.

```
<SwitchB> system-view
[SwitchB] ipv6
```

Configure a link aggregation group. Disable STP on the port before adding it into the link aggregation group.

```
[SwitchB] link-aggregation group 1 mode manual
[SwitchB] link-aggregation group 1 service-type tunnel
[SwitchB] interface GigabitEthernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] stp disable
[SwitchB-GigabitEthernet1/0/1] port link-aggregation group 1
[SwitchB-GigabitEthernet1/0/1] quit
```

Configure an IPv4 address for VLAN-interface 100.

```
[SwitchB] vlan 100
[SwitchB-vlan100] port GigabitEthernet 1/0/2
```

```
[SwitchB-vlan100] quit
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] ip address 5.1.1.1 24
[SwitchB-Vlan-interface100] quit
```

Configure a route to VLAN-interface 100 of Switch A. (Here the next-hop address of the static route is represented by [nexthop]. In practice, you should configure the real next-hop address according to the network.)

```
[SwitchB] ip route-static 2.1.1.1 24 [nexthop]
```

Configure an IPv6 address for VLAN-interface 101.

```
[SwitchB] vlan 101
[SwitchB-vlan101] port GigabitEthernet 1/0/3
[SwitchB-vlan101] quit
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] ipv6 address 2002:0501:0101:1::1/64
[SwitchB-Vlan-interface101] quit
```

Configure the 6to4 tunnel.

```
[SwitchB] interface tunnel 0
[SwitchB-Tunnel0] ipv6 address 2002:0501:0101::1/64
[SwitchB-Tunnel0] source vlan-interface 100
[SwitchB-Tunnel0] tunnel-protocol ipv6-ipv4 6to4
[SwitchB-Tunnel0] quit
```

Configure the tunnel to reference link aggregation group 1 in tunnel interface view.

```
[SwitchB-Tunnel0] aggregation-group 1
[SwitchB-Tunnel0] quit
```

Configure a static route whose destination address is 2002::/16 and the next hop is the tunnel interface.

```
[SwitchB] ipv6 route-static 2002:: 16 tunnel 0
```

Configuration verification

After the above configuration, ping Host B from Host A or ping Host A from Host B.

```
D:\>ping6 -s 2002:201:101:1::2 2002:501:101:1::2
```

```
Pinging 2002:501:101:1::2
from 2002:201:101:1::2 with 32 bytes of data:
```

```
Reply from 2002:501:101:1::2: bytes=32 time=13ms
Reply from 2002:501:101:1::2: bytes=32 time=1ms
Reply from 2002:501:101:1::2: bytes=32 time=1ms
Reply from 2002:501:101:1::2: bytes=32 time<1ms
```

```
Ping statistics for 2002:501:101:1::2:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
Approximate round trip times in milli-seconds:
    Minimum = 0ms, Maximum = 13ms, Average = 3ms
```

Configuring ISATAP Tunnel

Configuration Prerequisites IP addresses are configured for interfaces such as VLAN interface, and loopback interface on the device. Such an interface can serve as the source interface of a tunnel to ensure that the tunnel destination address is reachable.

Configuration Procedure Follow these steps to configure an ISATAP tunnel:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable IPv6	ipv6	Required By default, the IPv6 forwarding function is disabled.
Create a tunnel interface and enter tunnel interface view	interface tunnel <i>number</i>	Required By default, there is no tunnel interface on the device.
Configure an IPv6 address for the tunnel interface	Configure an IPv6 global unicast address or site-local address ipv6 address { <i>ipv6-address</i> <i>prefix-length</i> <i>ipv6-address/prefix-length</i> } ipv6 address <i>ipv6-address/prefix-length</i> eui-64	Required. Use either command. By default, no IPv6 global unicast address is configured for the tunnel interface.
	Configure an IPv6 link-local address ipv6 address auto link-local ipv6 address <i>ipv6-address</i> link-local	Optional By default, a link-local address will automatically be generated when an IPv6 global unicast address or link-local address is configured.
Disable the RA message suppression	undo ipv6 nd ra halt	Required Enabled by default.
Set an ISATAP tunnel	tunnel-protocol ipv6-ipv4 isatap	Required By default, the tunnel mode is manual. The same tunnel type should be configured at both ends of the tunnel. Otherwise, packet delivery will fail.
Configure a source address or interface for the tunnel	source { <i>ip-address</i> <i>interface-type</i> <i>interface-number</i> }	Required By default, no source address or interface is configured for the tunnel.
Reference a link aggregation group	aggregation-group <i>aggregation-group-id</i>	Required By default, no link aggregation group ID is referenced.

**CAUTION:**

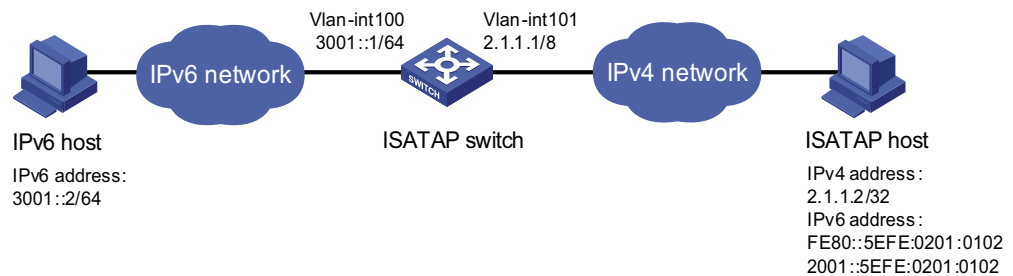
- If the addresses of the tunnel interfaces at the two ends of a tunnel are not in the same network segment, a forwarding route through the tunnel to the peer must be configured so that the encapsulated packet can be forwarded normally. You can configure static or dynamic routes at both ends of the tunnel.
- The automatic tunnel interfaces encapsulated with the same protocol cannot share the same source IP address.
- Automatic tunnels do not support dynamic routing.
- When you configure a static route, you need to configure a route to the destination address (the destination IP address of the packet, instead of the IPv4 address of the tunnel destination) and set the next-hop to the tunnel interface number or network address at the local end of the tunnel. Such a route must be configured at both ends of the tunnel.
- Before referencing a link aggregation group on the tunnel interface to receive and send packets, make sure that the aggregation group has been configured. Otherwise, the tunnel interface will not be up to communicate.

Configuration Example Network requirements

The destination address of a tunnel is an ISATAP address. It is required that IPv6 hosts in the IPv4 network can access the IPv6 network via an ISATAP tunnel.

Network diagram

Figure 161 Network diagram for an ISATAP tunnel

**Configuration procedure**

- Configuration on the switch

Enable IPv6.

```
<Switch> system-view
[Switch] ipv6
```

Configure a link aggregation group. Disable STP on the port before adding it into the link aggregation group.

```
[Switch] link-aggregation group 1 mode manual
[Switch] link-aggregation group 1 service-type tunnel
[Switch] interface GigabitEthernet 1/0/1
[Switch-GigabitEthernet1/0/1] stp disable
[Switch-GigabitEthernet1/0/1] port link-aggregation group 1
[Switch-GigabitEthernet1/0/1] quit
```

Configure addresses for interfaces.

```
[Switch] vlan 100
[Switch-vlan100] port GigabitEthernet 1/0/2
[Switch-vlan100] quit
[Switch] interface vlan-interface 100
[Switch-Vlan-interface100] ipv6 address 3001::1/64
[Switch-Vlan-interface100] quit
[Switch] vlan 101
[Switch-vlan101] port GigabitEthernet 1/0/3
[Switch-vlan101] quit
[Switch] interface vlan-interface 101
[Switch-Vlan-interface101] ip address 2.1.1.1 255.0.0.0
[Switch-Vlan-interface101] quit
```

Configure an ISATAP tunnel.

```
[Switch] interface tunnel 0
[Switch-Tunnel0] ipv6 address 2001::1/64 eui-64
[Switch-Tunnel0] source vlan-interface 101
[Switch-Tunnel0] tunnel-protocol ipv6-ipv4 isatap
```

Configure the tunnel to reference link aggregation group 1 in tunnel interface view.

```
[Switch-Tunnel0] aggregation-group 1
```

Disable the RA suppression so that hosts can acquire information such as the address prefix from the RA message released by the ISATAP switch.

```
[Switch-Tunnel0] undo ipv6 nd ra halt
```

■ Configuration on the ISATAP host

The specific configuration on the ISATAP host is related to its operating system. The following example shows the configuration of the host running the Windows XP.

On a Windows XP-based host, the ISATAP interface is usually interface 2. Configure the IPv4 address of the ISATAP router on the interface to complete the configuration on the host. Before doing that, display the ISATAP interface information:

```
C:\>ipv6 if 2
Interface 2: Automatic Tunneling Pseudo-Interface
  Guid {48FCE3FC-EC30-E50E-F1A7-71172AEEE3AE}
  does not use Neighbor Discovery
  does not use Router Discovery
  routing preference 1
  EUI-64 embedded IPv4 address: 0.0.0.0
  router link-layer address: 0.0.0.0
    preferred link-local fe80::5efe:2.1.1.2, life infinite
  link MTU 1280 (true link MTU 65515)
  current hop limit 128
  reachable time 42500ms (base 30000ms)
  retransmission interval 1000ms
```

```
DAD transmits 0
default site prefix length 48
```

A link-local address (fe80::5efe:2.1.1.2) in the ISATAP format was automatically generated for the ISATAP interface. Configure the IPv4 address of the ISATAP switch on the ISATAP interface.

```
C:\>ipv6 rlu 2 2.1.1.1
```

After carrying out the above command, look at the information on the ISATAP interface.

```
C:\>ipv6 if 2
Interface 2: Automatic Tunneling Pseudo-Interface
  Guid {48FCE3FC-EC30-E50E-F1A7-71172AEEE3AE}
  does not use Neighbor Discovery
  uses Router Discovery
  routing preference 1
  EUI-64 embedded IPv4 address: 2.1.1.2
  router link-layer address: 2.1.1.1
    preferred global 2001::5efe:2.1.1.2, life 29d23h59m46s/6d23h59m46s (public)
    preferred link-local fe80::5efe:2.1.1.2, life infinite
  link MTU 1500 (true link MTU 65515)
  current hop limit 255
  reachable time 42500ms (base 30000ms)
  retransmission interval 1000ms
  DAD transmits 0
  default site prefix length 48
```

By comparison, it is found that the host acquires the address prefix 2001::/64 and automatically generates the address 2001::5efe:2.1.1.2. Meanwhile, “uses Switch Discovery” is displayed, indicating that the switch discovery function is enabled on the host. At this time, ping the IPv6 address of the tunnel interface of the switch. If the address is successfully pinged, an ISATAP tunnel is established.

Configuration verification

After the above configurations, the ISATAP host can access the host in the IPV6 network.

Displaying and Maintaining Tunneling Configuration

To do...	Use the command...	Remarks
Display information about a specified tunnel interface	display interface tunnel [<i>number</i>]	Available in any view
Display IPv6 information related to a specified tunnel interface	display ipv6 interface tunnel <i>number</i>	Available in any view

Troubleshooting Tunneling Configuration

Symptom: After the configuration of related parameters such as tunnel source address, tunnel destination address, and tunnel type, the tunnel interface is still not up.

Solution: Follow the steps below:

- 1 The common cause is that the physical interface of the tunnel source is not up. Use the **display interface tunnel** or **display ipv6 interface tunnel** commands to view whether the physical interface of the tunnel source is up. If the physical

interface is down, use the **debugging tunnel event** command in user view to view the cause.

- 2 Another possible cause is that the tunnel destination is unreachable. Use the **display ipv6 routing-table** or **display ip routing-table** command to view whether the tunnel destination is reachable. If no routing entry is available for tunnel communication in the routing table, configure related routes.



This manual chiefly focuses on the IP multicast technology and device operations. Unless otherwise stated, the term "multicast" in this document refers to IP multicast.

Introduction to Multicast

As a technique coexisting with unicast and broadcast, the multicast technique effectively addresses the issue of point-to-multipoint data transmission. By allowing high-efficiency point-to-multipoint data transmission over a network, multicast greatly saves network bandwidth and reduces network load.

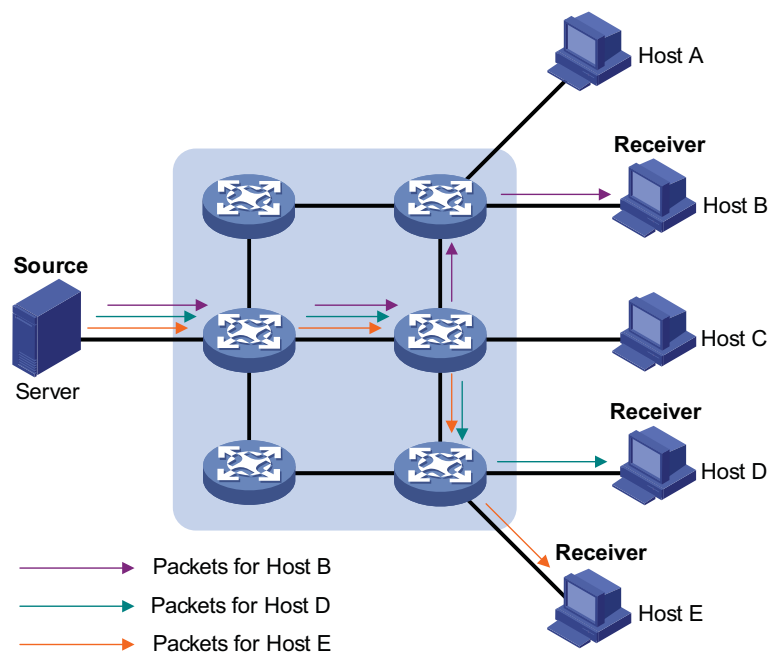
With the multicast technology, a network operator can easily provide new value-added services, such as live Webcasting, Web TV, distance learning, telemedicine, Web radio, real-time videoconferencing, and other bandwidth- and time-critical information services.

Comparison of Information Transmission Techniques

Unicast

In unicast, the information source sends a separate copy of information to each host that needs the information, as shown in Figure 162.

Figure 162 Unicast transmission



Assume that Hosts B, D and E need this information. The information source establishes a separate transmission channel for each of these hosts.

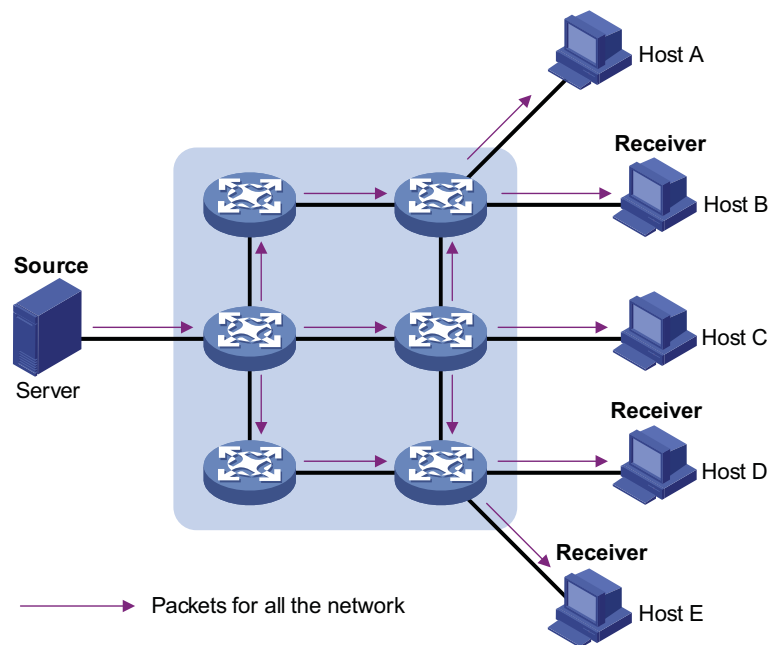
In unicast transmission, the traffic over the network is proportional to the number of hosts that need the information. If a large number of users need the information, the information source needs to send a copy of the same information to each of these users. This means a tremendous pressure on the information source and the network bandwidth.

As we can see from the information transmission process, unicast is not suitable for batch transmission of information.

Broadcast

In broadcast, the information source sends information to all hosts on the network, even if some hosts do not need the information, as shown in Figure 163.

Figure 163 Broadcast transmission



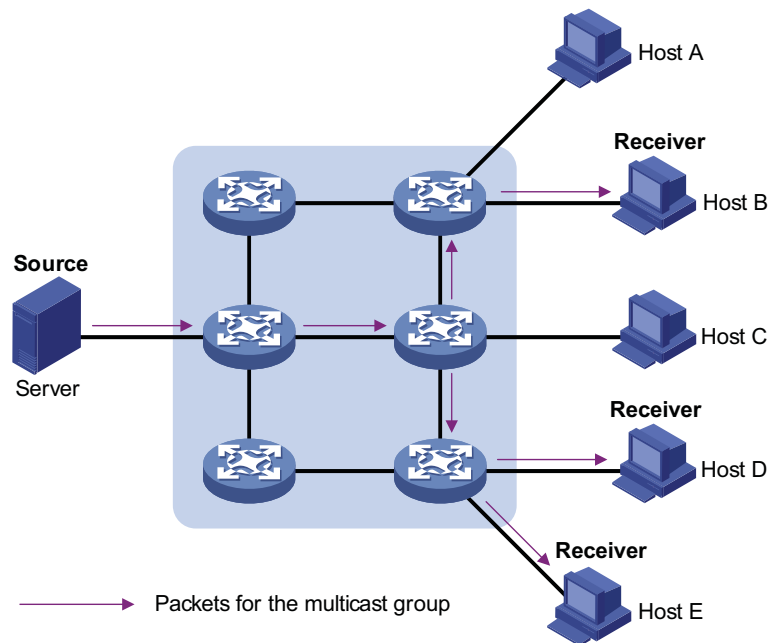
Assume that only Hosts B, D, and E need the information. If the information source broadcasts the information, Hosts A and C also receive it. In addition to information security issues, this also causes traffic flooding on the same network.

Therefore, broadcast is disadvantageous in transmitting data to specific hosts; moreover, broadcast transmission is a significant usage of network resources.

Multicast

As discussed above, the unicast and broadcast techniques are unable to provide point-to-multipoint data transmissions with the minimum network consumption.

The multicast technique has solved this problem. When some hosts on the network need multicast information, the multicast source (Source in the figure) sends only one copy of the information. Multicast distribution trees are built for the multicast packets through multicast routing protocols, and the packets are replicated only on nodes where the trees branch, as shown in Figure 164:

Figure 164 Multicast transmission

Assume that Hosts B, D and E need the information. To receive the information correctly, these hosts need to join a receiver set, which is known as a multicast group. The routers on the network duplicate and forward the information based on the distribution of the receivers in this set. Finally, the information is correctly delivered to Hosts B, D, and E.

To sum up, multicast has the following advantages:

- Over unicast: As multicast traffic flows to the node the farthest possible from the source before it is replicated and distributed, an increase of the number of hosts will not remarkably add to the network load.
- Over broadcast: As multicast data is sent only to the receivers that need it, multicast uses the network bandwidth reasonably and brings no waste of network resources, and enhances network security.

Roles in Multicast The following roles are involved in multicast transmission:

- An information sender is referred to as a Multicast Source ("Source" in Figure 164).
- Each receiver is a Multicast Group Member ("Receiver" in Figure 164).
- All receivers interested in the same information form a Multicast Group. Multicast groups are not subject to geographic restrictions.
- A router that supports Layer 3 multicast is called multicast router or Layer 3 multicast device. In addition to providing the multicast routing function, a multicast router can also manage multicast group members.

For a better understanding of the multicast concept, you can assimilate multicast transmission to the transmission of TV programs, as shown in Table 51.

Table 51 An analogy between TV transmission and multicast transmission

Step	TV transmission	Multicast transmission
1	A TV station transmits a TV program through a channel.	A multicast source sends multicast data to a multicast group.
2	A user tunes the TV set to the channel.	A receiver joins the multicast group.
3	The user starts to watch the TV program transmitted by the TV station via the channel.	The receiver starts to receive the multicast data that the source sends to the multicast group.
4	The user turns off the TV set or tunes to another channel.	The receiver leaves the multicast group or joins another group.



- *A multicast source does not necessarily belong to a multicast group. Namely, a multicast source is not necessarily a multicast data receiver.*
- *A multicast source can send data to multiple multicast groups at the same time, and multiple multicast sources can send data to the same multicast group at the same time.*

Advantages and Applications of Multicast

Advantages of multicast

Advantages of the multicast technique include:

- Enhanced efficiency: reduces the CPU load of information source servers and network devices.
- Optimal performance: reduces redundant traffic.
- Distributive application: Enables point-to-multiple-point applications at the price of the minimum network resources.

Applications of multicast

Applications of the multicast technique include:

- Multimedia and streaming applications, such as Web TV, Web radio, and real-time video/audio conferencing.
- Communication for training and cooperative operations, such as distance learning and telemedicine.
- Data warehouse and financial applications (stock quotes).
- Any other point-to-multiple-point data distribution application.

Multicast Models

Based on how the receivers treat the multicast sources, there are two multicast models:

ASM model

In the ASM model, any sender can send information to a multicast group as a multicast source, and numbers of receivers can join a multicast group identified by a group address and obtain multicast information addressed to that multicast group. In this model, receivers are not aware of the position of multicast sources in advance. However, they can join or leave the multicast group at any time.

SSM model

In the practical life, users may be interested in the multicast data from only certain multicast sources. The SSM model provides a transmission service that allows users to specify the multicast sources they are interested in at the client side.

The radical difference between the SSM model and the ASM model is that in the SSM model, receivers already know the locations of the multicast sources by some other means. In addition, the SSM model uses a multicast address range that is different from that of the ASM model, and dedicated multicast forwarding paths are established between receivers and the specified multicast sources.

Multicast Architecture

IP multicast addresses the following questions:

- Where should the multicast source transmit information to? (multicast addressing)
- What receivers exist on the network? (host registration)
- Where is the multicast source from which the receivers need to receive multicast data? (multicast source discovery)
- How should information be transmitted to the receivers? (multicast routing)

IP multicast falls in the scope of end-to-end service. The multicast architecture involves the following four parts:

- 1 Addressing mechanism: Information is sent from a multicast source to a group of receivers through a multicast address.
- 2 Host registration: Receiver hosts are allowed to join and leave multicast groups dynamically. This mechanism is the basis for group membership management.
- 3 Multicast routing: A multicast distribution tree (namely a forwarding path tree for multicast data on the network) is constructed for delivering multicast data from a multicast source to receivers.
- 4 Multicast applications: A software system that supports multicast applications, such as video conferencing, must be installed on multicast sources and receiver hosts, and the TCP/IP stack must support reception and transmission of multicast data.

Multicast Addresses

To allow communication between multicast sources and multicast group members, network-layer multicast addresses, namely, multicast IP addresses must be provided. In addition, a technique must be available to map multicast IP addresses to link-layer multicast MAC addresses.

IPv4 multicast addresses

Internet Assigned Numbers Authority (IANA) assigned the Class D address space (224.0.0.0 to 239.255.255.255) for IPv4 multicast. The specific address blocks and usages are shown in Table 52.

Table 52 Class D IP address blocks and description

Address block	Description
224.0.0.0 to 224.0.0.255	Reserved permanent group addresses. The IP address 224.0.0.0 is reserved, and other IP addresses can be used by routing protocols and for topology searching, protocol maintenance, and so on. Commonly used permanent group addresses are listed in Table 53. A packet destined for an address in this block will not be forwarded beyond the local subnet regardless of the Time to Live (TTL) value in the IP header.
224.0.1.0 to 238.255.255.255	Globally scoped group addresses. This block includes two types of designated group addresses: <ul style="list-style-type: none"> ■ 232.0.0.0/8: SSM group addresses, and ■ 233.0.0.0/8: Glop group addresses; for details, see RFC 2770.
239.0.0.0 to 239.255.255.255	Administratively scoped multicast addresses. These addresses are considered to be locally rather than globally unique, and can be reused in domains administered by different organizations without causing conflicts. For details, refer to RFC 2365.



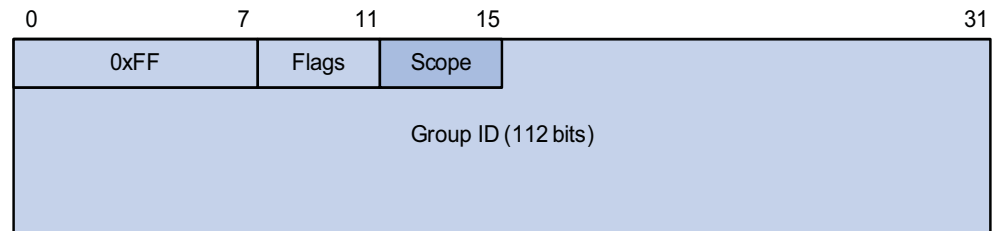
- *The membership of a group is dynamic. Hosts can join or leave multicast groups at any time.*
- *Glop" is a mechanism for assigning multicast addresses between different autonomous systems (ASs). By filling an AS number into the middle two bytes of 233.0.0.0, you get 255 multicast addresses for that AS.*

Table 53 Some reserved multicast addresses

Address	Description
224.0.0.1	All systems on this subnet, including hosts and routers
224.0.0.2	All multicast routers on this subnet
224.0.0.3	Unassigned
224.0.0.4	Distance Vector Multicast Routing Protocol (DVMRP) routers
224.0.0.5	Open Shortest Path First (OSPF) routers
224.0.0.6	OSPF designated routers/backup designated routers
224.0.0.7	Shared Tree (ST) routers
224.0.0.8	ST hosts
224.0.0.9	Routing Information Protocol version 2 (RIPv2) routers
224.0.0.11	Mobile agents
224.0.0.12	Dynamic Host Configuration Protocol (DHCP) server/relay agent
224.0.0.13	All Protocol Independent Multicast (PIM) routers
224.0.0.14	Resource Reservation Protocol (RSVP) encapsulation
224.0.0.15	All Core-Based Tree (CBT) routers
224.0.0.16	Designated Subnetwork Bandwidth Management (SBM)
224.0.0.17	All SBMs
224.0.0.18	Virtual Router Redundancy Protocol (VRRP)

IPv6 Multicast Addresses

As defined in RFC 4291, the format of an IPv6 multicast is as follows:

Figure 165 IPv6 multicast format

- 0xFF: 8 bits, indicating that this address is an IPv6 multicast address.
- Flags: 4 bits, of which the high-order flag is reserved and set to 0; the definition and usage of the second bit can be found in RFC 3956; and definition and usage of the third bit can be found in RFC 3306; the low-order bit is the Transient (T) flag. When set to 0, the T flag indicates a permanently-assigned multicast address assigned by IANA; when set to 1, the T flag indicates a transient, or dynamically assigned multicast address.
- Scope: 4 bits, indicating the scope of the IPv6 internetwork for which the multicast traffic is intended. Possible values of this field are given in Table 54.
- Reserved: 80 bits, all set to 0 currently.
- Group ID: 112 bits, identifying the multicast group. For details about this field, refer to RFC 3306.

Table 54 Values of the Scope field

Value	Meaning
0, 3, F	Reserved
1	Node-local scope
2	Link-local scope
4	Admin-local scope
5	Site-local scope
6, 7, 9 through D	Unassigned
8	Organization-local scope
E	Global scope

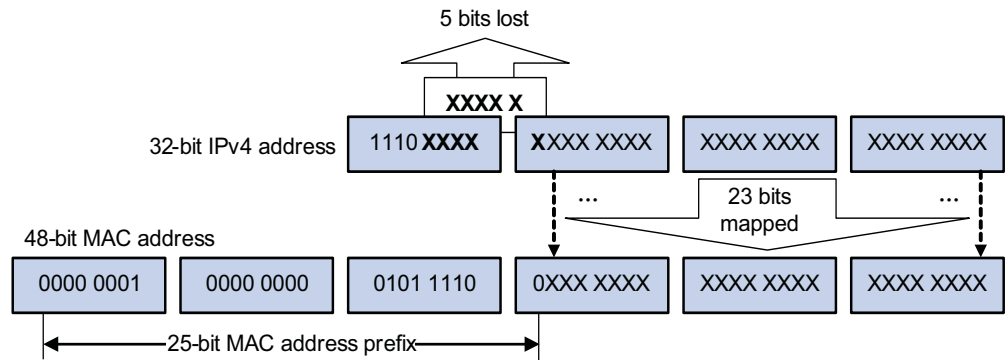
Ethernet multicast MAC addresses

When a unicast IP packet is transmitted over Ethernet, the destination MAC address is the MAC address of the receiver. When a multicast packet is transmitted over Ethernet, however, the destination address is a multicast MAC address because the packet is directed to a group formed by a number of receivers, rather than to one specific receiver.

1 IPv4 multicast MAC addresses

As defined by IANA, the high-order 24 bits of an IPv4 multicast MAC address are 0x01005e, bit 25 is 0x0, and the low-order 23 bits are the low-order 23 bits of a multicast IPv4 address. The IPv4-to-MAC mapping relation is shown in Figure 166.

Figure 166 IPv4-to-MAC address mapping

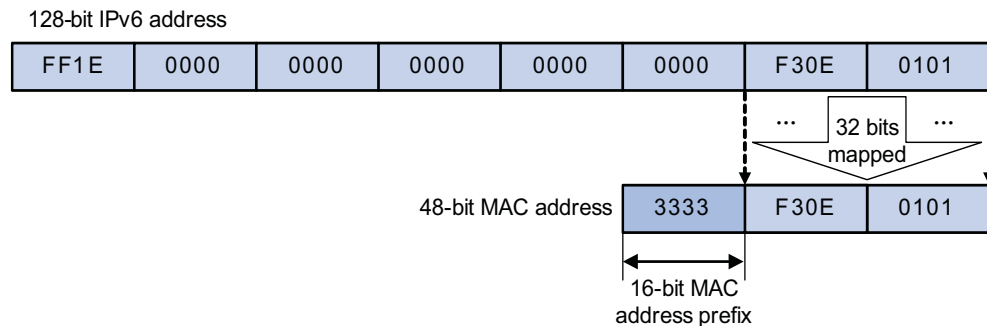


The high-order four bits of a multicast IPv4 address are 1110, indicating that this address is a multicast address, and only 23 bits of the remaining 28 bits are mapped to a MAC address, so five bits of the multicast IPv4 address are lost. As a result, 32 multicast IPv4 addresses map to the same MAC address. Therefore, in Layer 2 multicast forwarding, a device may receive some multicast data addressed for other IPv4 multicast groups, and such redundant data needs to be filtered by the upper layer.

2 IPv6 multicast MAC addresses

The high-order 16 bits of an IPv6 multicast MAC address are 0x3333, and the low-order 32 bits are the low-order 32 bits of a multicast IPv6 address. Figure 167 shows an example of mapping an IPv6 multicast address, FF1E::F30E:0101, to a MAC address.

Figure 167 An example of IPv6-to-MAC address mapping



Multicast Protocols



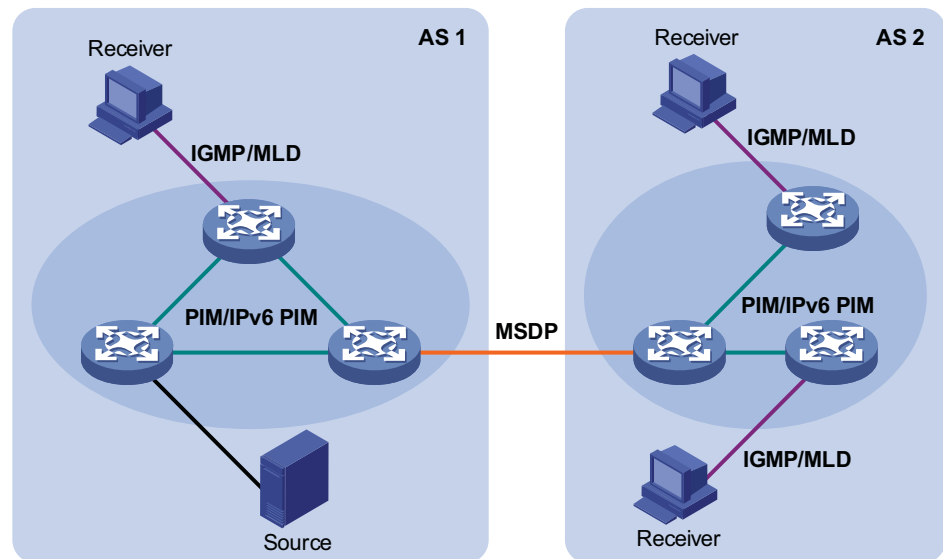
- Generally, we refer to IP multicast working at the network layer as Layer 3 multicast and the corresponding multicast protocols as Layer 3 multicast protocols, which include IGMP/MLD, PIM/IPv6 PIM, and MSDP; we refer to IP multicast working at the data link layer as Layer 2 multicast and the corresponding multicast protocols as Layer 2 multicast protocols, which include IGMP Snooping/MLD Snooping, and multicast VLAN/IPv6 multicast VLAN.
- IGMP Snooping, IGMP, multicast VLAN, PIM and MSDP are for IPv4, MLD Snooping, MLD, IPv6 multicast VLAN, and IPv6 PIM are for IPv6.

This section provides only general descriptions about applications and functions of the Layer 2 and Layer 3 multicast protocols in a network. For details of these protocols, refer to the respective chapters.

Layer 3 multicast protocols

Layer 3 multicast protocols include multicast group management protocols and multicast routing protocols. Figure 168 describes where these multicast protocols are in a network.

Figure 168 Positions of Layer 3 multicast protocols



1 Multicast management protocols

Typically, the internet group management protocol (IGMP) or multicast listener discovery protocol (MLD) is used between hosts and Layer 3 multicast devices directly connected with the hosts. These protocols define the mechanism of establishing and maintaining group memberships between hosts and Layer 3 multicast devices.

2 Multicast routing protocols

A multicast routing protocol runs on Layer 3 multicast devices to establish and maintain multicast routes and forward multicast packets correctly and efficiently. Multicast routes constitute a loop-free data transmission path from a data source to multiple receivers, namely, a multicast distribution tree.

In the ASM model, multicast routes come in intra-domain routes and inter-domain routes.

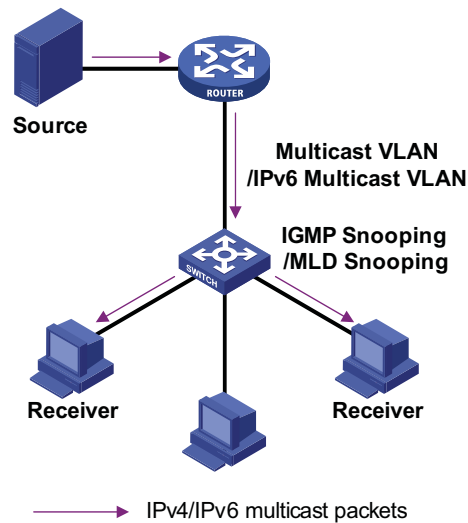
- An intra-domain multicast routing protocol is used to discover multicast sources and build multicast distribution trees within an AS so as to deliver multicast data to receivers. Among a variety of mature intra-domain multicast routing protocols, protocol independent multicast (PIM) is a popular one. Based on the forwarding mechanism, PIM comes in two modes - dense mode (often referred to as PIM-DM) and sparse mode (often referred to as PIM-SM).
- An inter-domain multicast routing protocol is used for delivery of multicast information between two ASs. So far, mature solutions include multicast source discovery protocol (MSDP).

For the SSM model, multicast routes are not divided into inter-domain routes and intra-domain routes. Since receivers know the position of the multicast source, channels established through PIM-SM are sufficient for multicast information transport.

Layer 2 multicast protocols

Layer 2 multicast protocols include IGMP Snooping/MLD Snooping and multicast VLAN/IPv6 multicast VLAN. Figure 169 shows where these protocols are in the network.

Figure 169 Position of Layer 2 multicast protocols



1 IGMP Snooping/MLD Snooping

Running on Layer 2 devices, Internet Group Management Protocol Snooping (IGMP Snooping) and Multicast Listener Discovery Snooping (MLD Snooping) are multicast constraining mechanisms that manage and control multicast groups by listening to and analyzing IGMP or MLD messages exchanged between the hosts and Layer 3 multicast devices, thus effectively controlling the flooding of multicast data in a Layer 2 network.

2 Multicast VLAN/IPv6 multicast VLAN

In the traditional multicast-on-demand mode, when users in different VLANs on a Layer 2 device need multicast information, the upstream Layer 3 device needs to forward a separate copy of the multicast data to each VLAN of the Layer 2 device. With the multicast VLAN or IPv6 multicast VLAN feature enabled on the Layer 2 device, the Layer 3 multicast device needs to send only one copy of multicast to the multicast VLAN or IPv6 multicast VLAN on the Layer 2 device. This avoids waste of network bandwidth and extra burden on the Layer 3 device.

Multicast Packet Forwarding Mechanism

In a multicast model, a multicast source sends information to the host group identified by the multicast group address in the destination address field of IP multicast packets. Therefore, to deliver multicast packets to receivers located in different parts of the network, multicast routers on the forwarding path usually need to forward multicast packets received on one incoming interface to multiple outgoing interfaces. Compared with a unicast model, a multicast model is more complex in the following aspects.

- To ensure multicast packet transmission in the network, unicast routing tables or multicast routing tables specially provided for multicast must be used as guidance for multicast forwarding.
- To process the same multicast information from different peers received on different interfaces of the same device, every multicast packet is subject to a reverse path forwarding (RPF) check on the incoming interface. The result of the RPF check determines whether the packet will be forwarded or discarded. The RPF check mechanism is the basis for most multicast routing protocols to implement multicast forwarding.



For details about the RPF mechanism, refer to “RPF Mechanism” on page 701.

When configuring IGMP Snooping, go to the following sections for information you are interested in:

- “IGMP Snooping Overview” on page 553
- “IGMP Snooping Configuration Task List” on page 558
- “Displaying and Maintaining IGMP Snooping” on page 569
- “IGMP Snooping Configuration Examples” on page 570
- “Troubleshooting IGMP Snooping Configuration” on page 577

IGMP Snooping Overview

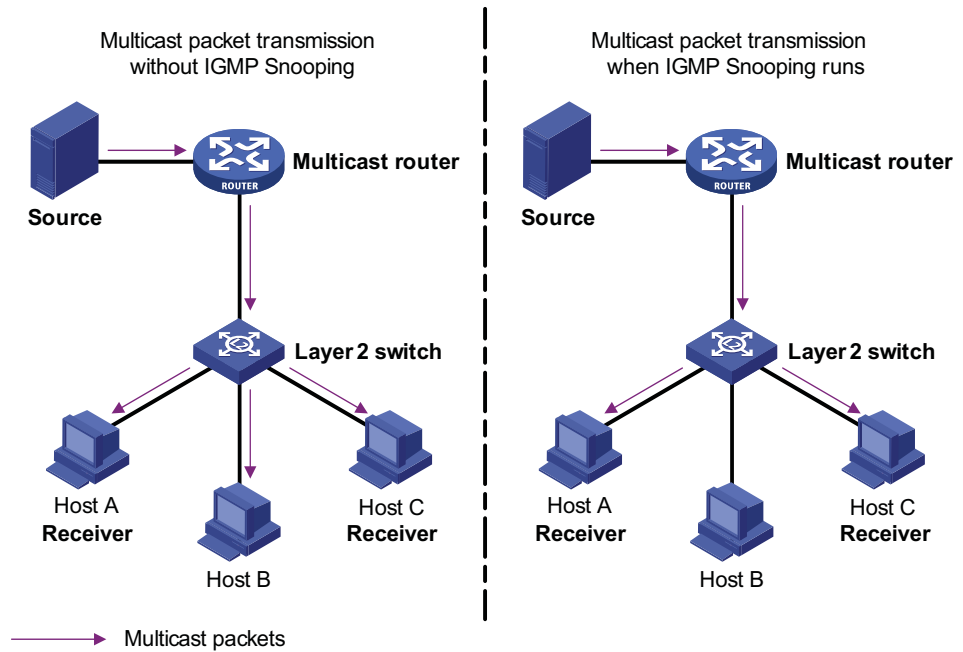
Internet Group Management Protocol Snooping (IGMP Snooping) is a multicast constraining mechanism that runs on Layer 2 devices to manage and control multicast groups.

Principle of IGMP Snooping

By analyzing received IGMP messages, a Layer 2 device running IGMP Snooping establishes mappings between ports and multicast IP addresses and forwards multicast data based on these mappings.

As shown in Figure 170, when IGMP Snooping is not running on the switch, multicast packets are broadcast to all devices at Layer 2. When IGMP Snooping is running on the switch, multicast packets for known multicast groups are multicast to the receivers, rather than broadcast to all hosts, at Layer 2.

Figure 170 Before and after IGMP Snooping is enabled on the Layer 2 device

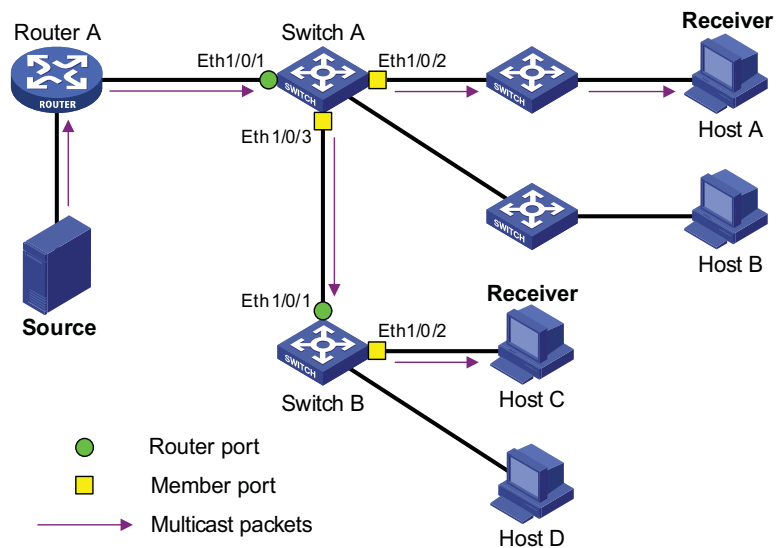


Basic Concepts in IGMP Snooping

IGMP Snooping related ports

As shown in Figure 171, Router A connects to the multicast source, IGMP Snooping runs on Switch A and Switch B, Host A and Host C are receiver hosts (namely, multicast group members).

Figure 171 IGMP Snooping related ports



Ports involved in IGMP Snooping, as shown in Figure 171, are described as follows:

- Router port: A router port is a port on the Ethernet switch that leads switch towards the Layer 3 multicast device (DR or IGMP querier). In the figure, Ethernet 1/0/1 of Switch A and Ethernet 1/0/1 of Switch B are router ports. The

switch registers all its local router ports (including static and dynamic router ports) in its router port list.

- Member port: A member port is a port on the Ethernet switch that leads switch towards multicast group members. In the figure, Ethernet 1/0/2 and Ethernet 1/0/3 of Switch A and Ethernet 1/0/2 of Switch B are member ports. The switch registers all the member ports (including static and dynamic member ports) on the local device in its IGMP Snooping forwarding table.



- *Whenever mentioned in this document, a router port is a port on the switch that leads the switch to a Layer 3 multicast device, rather than a port on a router.*
- *An IGMP-snooping-enabled switch deems that all its ports on which IGMP general queries with the source address other than 0.0.0.0 or PIM hello messages are received to be router ports.*

Aging timers for dynamic ports in IGMP Snooping and related messages and actions

Table 55 Aging timers for dynamic ports in IGMP Snooping and related messages and actions

Timer	Description	Message before expiry	Action after expiry
Router port aging timer	For each router port, the switch sets a timer initialized to the aging time of the route port.	IGMP general query of which the source address is not 0.0.0.0 or PIM hello	The switch removes this port from its router port list.
Member port aging timer	When a port joins a multicast group, the switch sets a timer for the port, which is initialized to the member port aging time.	IGMP membership report	The switch removes this port from the multicast group forwarding table.



The port aging mechanism of IGMP Snooping works only for dynamic ports; a static port will never age out.

Work Mechanism of IGMP Snooping

A switch running IGMP Snooping performs different actions when it receives different IGMP messages, as follows:

When receiving a general query

The IGMP querier periodically sends IGMP general queries to all hosts and routers (224.0.0.1) on the local subnet to find out whether active multicast group members exist on the subnet.

Upon receiving an IGMP general query, the switch forwards it through all ports in the VLAN except the receiving port and performs the following to the receiving port:

- If the receiving port is a router port existing in its router port list, the switch resets the aging timer of this router port.
- If the receiving port is not a router port existing in its router port list, the switch adds it into its router port list and sets an aging timer for this router port.

When receiving a membership report

A host sends an IGMP report to the multicast router in the following circumstances:

- Upon receiving an IGMP query, a multicast group member host responds with an IGMP report.
- When intended to join a multicast group, a host sends an IGMP report to the multicast router to announce that it is interested in the multicast information addressed to that group.

Upon receiving an IGMP report, the switch forwards it through all the router ports in the VLAN, resolves the address of the reported multicast group, and performs the following:

- If no forwarding table entry exists for the reported group, the switch creates an entry, adds the port as member port to the outgoing port list, and starts a member port aging timer for that port.
- If a forwarding table entry exists for the reported group, but the port is not included in the outgoing port list for that group, the switch adds the port as a member port to the outgoing port list, and starts a member port aging timer for that port.
- If a forwarding table entry exists for the reported group and the port is included in the outgoing port list, which means that this port is already a member port, the switch resets the member port aging timer for that port.



A switch does not forward an IGMP report through a non-router port. The reason is as follows: Due to the IGMP report suppression mechanism, if the switch forwards a report message through a member port, all the attached hosts listening to the reported multicast address will suppress their own reports upon hearing this report, and this will prevent the switch from knowing whether any hosts attached to that port are still active members of the reported multicast group.

For the description of IGMP report suppression mechanism, refer to “Work Mechanism of IGMPv1” on page 613.

When receiving a leave group message

When an IGMPv1 host leaves a multicast group, the host does not send an IGMP leave group message, so the switch cannot know immediately that the host has left the multicast group. However, as the host stops sending IGMP reports as soon as it leaves a multicast group, the switch deletes the forwarding entry for the member port corresponding to the host from the forwarding table when its aging timer expires.

When an IGMPv2 or IGMPv3 host leaves a multicast group, the host sends an IGMP leave group message to the multicast router.

When the switch hears a group-specific IGMP leave group message on a member port, it first checks whether a forwarding table entry for that group exists, and, if one exists, whether its outgoing port list contains that port.

- If the forwarding table entry does not exist or if its outgoing port list does not contain the port, the switch discards the IGMP leave group message instead of forwarding it to any port.

- If the forwarding table entry exists and its outgoing port list contains the port, the switch forwards the leave group message to all router ports in the VLAN. Because the switch does not know whether any other hosts attached to the port are still listening to that group address, the switch does not immediately remove the port from the outgoing port list of the forwarding table entry for that group; instead, it resets the member port aging timer for the port.

Upon receiving the IGMP leave group message from a host, the IGMP querier resolves from the message the address of the multicast group that the host just left and sends an IGMP group-specific query to that multicast group through the port that received the leave group message. Upon hearing the IGMP group-specific query, the switch forwards it through all its router ports in the VLAN and all member ports for that multicast group, and performs the following:

- If any IGMP report in response to the group-specific query is heard on a member port before its aging timer expires, this means that some host attached to the port is receiving or expecting to receive multicast data for that multicast group. The switch resets the aging timer of the member port.
- If no IGMP report in response to the group-specific query is heard on a member port before its aging timer expires, this means that no hosts attached to the port are still listening to that group address: the switch removes the port from the outgoing port list of the forwarding table entry for that multicast group when the aging timer expires.

Processing of Multicast Protocol Messages

With Layer 3 multicast routing enabled, an IGMP Snooping switch processes multicast protocol messages differently under different conditions, specifically as follows:

- 1 If only IGMP is enabled, or both IGMP and PIM are enabled on the switch, the switch handles multicast protocol messages in the normal way.
- 2 In only PIM is enabled on the switch:
 - The switch broadcasts IGMP messages as unknown messages in the VLAN.
 - Upon receiving a PIM hello message, the switch will maintain the corresponding router port.
- 3 When IGMP is disabled on the switch, or when IGMP forwarding entries are cleared (by using the **reset igmp group** command):
 - If PIM is disabled, the switch clears all its Layer 2 multicast entries and router ports.
 - If PIM is enabled, the switch clears only its Layer 2 multicast entries without deleting its router ports.
- 4 When PIM is disabled on the switch:
 - If IGMP is disabled, the switch clears all its router ports.
 - If IGMP is enabled, the switch maintains all its Layer 2 multicast entries and router ports.

Protocols and Standards

IGMP Snooping is documented in:

- RFC 4541: Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches

IGMP Snooping Configuration Task List

Complete these tasks to configure IGMP Snooping:

Task	Remarks	
"Configuring Basic Functions of IGMP Snooping" on page 559	"Enabling IGMP Snooping" on page 559	Required
	"Configuring the Version of IGMP Snooping" on page 559	Optional
"Configuring IGMP Snooping Port Functions" on page 560	"Configuring Aging Timers for Dynamic Ports" on page 560	Optional
	"Configuring Static Ports" on page 561	Optional
	"Configuring Simulated Joining" on page 561	Optional
	"Configuring Fast Leave Processing" on page 562	Optional
"Configuring IGMP Snooping Querier" on page 563	"Enabling IGMP Snooping Querier" on page 563	Optional
	"Configuring IGMP Queries and Responses" on page 564	Optional
	"Configuring Source IP Address of IGMP Queries" on page 565	Optional
"Configuring an IGMP Snooping Policy" on page 565	"Configuring a Multicast Group Filter" on page 566	Optional
	"Configuring Multicast Source Port Filtering" on page 566	Optional
	"Configuring the Function of Dropping Unknown Multicast Data" on page 567	Optional
	"Configuring IGMP Report Suppression" on page 568	Optional
	"Configuring Maximum Multicast Groups that Can Be Joined on a Port" on page 568	Optional
	"Configuring Multicast Group Replacement" on page 569	Optional



- *Configurations made in IGMP Snooping view are effective for all VLANs, while configurations made in VLAN view are effective only for ports belonging to the current VLAN. For a given VLAN, a configuration made in IGMP Snooping view is effective only if the same configuration is not made in VLAN view.*
- *Configurations made in IGMP Snooping view are effective for all ports; configurations made in Ethernet port view are effective only for the current port; configurations made in manual port group view are effective only for all the ports in the current port group; configurations made in aggregation group view are effective only for the master port. For a given port, a configuration made in IGMP Snooping view is effective only if the same configuration is not made in Ethernet port view or port group view.*

Configuring Basic Functions of IGMP Snooping

Configuration Prerequisites Before configuring the basic functions of IGMP Snooping, complete the following task:

- Configure the corresponding VLANs.

Before configuring the basic functions of IGMP Snooping, prepare the following data:

- Version of IGMP Snooping.

Enabling IGMP Snooping Follow these steps to enable IGMP Snooping:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable IGMP Snooping globally and enter IGMP-Snooping view	igmp-snooping	Required Disabled by default
Return to system view	quit	-
Enter VLAN view	vlan <i>vlan-id</i>	-
Enable IGMP Snooping in the VLAN	igmp-snooping enable	Required Disabled by default



- *IGMP Snooping must be enabled globally before it can be enabled in a VLAN.*
- *After enabling IGMP Snooping in a VLAN, you cannot enable IGMP and/or PIM on the corresponding VLAN interface, and vice versa.*
- *When you enable IGMP Snooping in a specified VLAN, this function takes effect for Ethernet ports in this VLAN only.*

Configuring the Version of IGMP Snooping

By configuring an IGMP Snooping version, you actually configure the version of IGMP messages that IGMP Snooping can process.

- IGMP Snooping version 2 can process IGMPv1 and IGMPv2 messages, but not IGMPv3 messages, which will be flooded in the VLAN.
- IGMP Snooping version 3 can process IGMPv1, IGMPv2 and IGMPv3 messages.

Follow these steps to configure the version of IGMP Snooping:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter VLAN view	vlan <i>vlan-id</i>	-
Configure the version of IGMP Snooping	igmp-snooping version <i>version-number</i>	Optional Version 2 by default



CAUTION: *If you switch IGMP Snooping from version 3 to version 2, the system will clear all IGMP Snooping forwarding entries from dynamic joins, and will*

- *Keep forwarding entries for version 3 static (*, G) joins;*
- *Clear forwarding entries from version 3 static (S, G) joins, which will be restored when IGMP Snooping is switched back to version 3.*

For details about static joins, Refer to “Configuring Static Ports” on page 561.

Configuring IGMP Snooping Port Functions

Configuration Prerequisites

Before configuring IGMP Snooping port functions, complete the following tasks:

- Enable IGMP Snooping in the VLAN or enable IGMP on the desired VLAN interface
- Configure the corresponding port groups.

Before configuring IGMP Snooping port functions, prepare the following data:

- Aging time of router ports,
- Aging timer of member ports, and
- Multicast group and multicast source addresses

Configuring Aging Timers for Dynamic Ports

If the switch receives no IGMP general queries or PIM hello messages on a dynamic router port, the switch removes the port from the router port list when the aging timer of the port expires.

If the switch receives no IGMP reports for a multicast group on a dynamic member port, the switch removes the port from the outgoing port list of the forwarding table entry for that multicast group when the aging timer of the port for that group expires.

If multicast group memberships change frequently, you can set a relatively small value for the member port aging timer, and vice versa.

Configuring aging timers for dynamic ports globally

Follow these steps to configure aging timers for dynamic ports globally:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter IGMP Snooping view	igmp-snooping	-
Configure router port aging time	router-aging-time <i>interval</i>	Optional 105 seconds by default
Configure member port aging time	host-aging-time <i>interval</i>	Optional 260 seconds by default

Configuring aging timers for dynamic ports in a VLAN

Follow these steps to configure aging timers for dynamic ports in a VLAN:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter VLAN view	vlan <i>vlan-id</i>	-
Configure router port aging time	igmp-snooping router-aging-time <i>interval</i>	Optional 105 seconds by default
Configure member port aging time	igmp-snooping host-aging-time <i>interval</i>	Optional 260 seconds by default

Configuring Static Ports

If all the hosts attached to a port are interested in the multicast data addressed to a particular multicast group or the multicast data that a particular multicast source sends to a particular group, you can configure static (*, G) or (S, G) joining on that port, namely configure the port as a group-specific or source-and-group-specific static member port.

You can configure a port of a switch to be a static router port, through which the switch can forward all the multicast traffic it received.

Follow these steps to configure static ports:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter the corresponding view	Enter Ethernet port view interface <i>interface-type</i> <i>interface-number</i>	Use either command.
	Enter port group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	
Configure the port(s) as static member port(s)	igmp-snooping static-group <i>group-address</i> [source-ip <i>source_address</i>] vlan <i>vlan-id</i>	Required Disabled by default
Configure the port(s) as static router port(s)	igmp-snooping static-router-port vlan <i>vlan-id</i>	Required Disabled by default



- The static (S, G) joining function is available only if a valid multicast source address is specified and IGMP Snooping version 3 is currently running on the switch.
- A static member port does not respond to queries from the IGMP querier; when static (*, G) or (S, G) joining is enabled or disabled on a port, the port does not send an unsolicited IGMP report or an IGMP leave group message.
- Static member ports and static router ports never age out. To remove such a port, you need to use the corresponding command.

Configuring Simulated Joining

Generally, a host running IGMP responds to IGMP queries from the IGMP querier. If a host fails to respond due to some reasons, the multicast router may deem that no member of this multicast group exists on the network segment, and therefore will remove the corresponding forwarding path.

To avoid this situation from happening, you can enable simulated joining on a port of the switch, namely configure the port as a simulated member host for a multicast group. When an IGMP query is heard, the simulated host gives a response. Thus, the switch can continue receiving multicast data.

A simulated host acts like a real host, as follows:

- When a port is configured as a simulated member host, the switch sends an unsolicited IGMP report through that port.
- After a port is configured as a simulated member host, the switch responds to IGMP general queries by sending IGMP reports through that port.
- When the simulated joining function is disabled on a port, the switch sends an IGMP leave group message through that port.

Follow these steps to configure simulated joining:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Enter the corresponding view	Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	Use either command
	Enter port group view	port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	
Configure simulated (*, G) or (S, G) joining		igmp-snooping host-join <i>group-address</i> [source-ip <i>source-address</i>] vlan <i>vlan-id</i>	Required Disabled by default



- *Each simulated host is equivalent to an independent host. For example, when receiving an IGMP query, the simulated host corresponding to each configuration responds respectively.*
- *Unlike a static member port, a port configured as a simulated member host will age out like a dynamic member port.*

Configuring Fast Leave Processing

The fast leave processing feature allows the switch to process IGMP leave group messages in a fast way. With the fast leave processing feature enabled, when receiving an IGMP leave group message on a port, the switch immediately removes that port from the outgoing port list of the forwarding table entry for the indicated group. Then, when receiving IGMP group-specific queries for that multicast group, the switch will not forward them to that port.

In VLANs where only one host is attached to each port, fast leave processing helps improve bandwidth and resource usage.

Configuring fast leave processing globally

Follow these steps to configure fast leave processing globally:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter IGMP Snooping view	igmp-snooping	-
Enable fast leave processing	fast-leave [vlan <i>vlan-list</i>]	Required Disabled by default

Configuring fast leave processing on a port or a group of ports

Follow these steps to configure fast leave processing on a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter the corresponding view	Enter Ethernet port view interface <i>interface-type</i> <i>interface-number</i>	Use either command
	Enter port group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	
Enable fast leave processing	igmp-snooping fast-leave [vlan <i>vlan-list</i>]	Required Disabled by default



CAUTION: If fast leave processing is enabled on a port to which more than one host is attached, when one host leaves a multicast group, the other hosts attached to the port and interested in the same multicast group will fail to receive multicast data for that group.

Configuring IGMP Snooping Querier

Configuration Prerequisites

Before configuring IGMP Snooping querier, complete the following task:

- Enable IGMP Snooping in the VLAN.

Before configuring IGMP Snooping querier, prepare the following data:

- IGMP general query interval,
- IGMP last-member query interval,
- Maximum response time to IGMP general queries,
- Source address of IGMP general queries, and
- Source address of IGMP group-specific queries.

Enabling IGMP Snooping Querier

In an IP multicast network running IGMP, a multicast router or Layer 3 multicast switch is responsible for sending IGMP general queries, so that all Layer 3 multicast devices can establish and maintain multicast forwarding entries, thus to forward multicast traffic correctly at the network layer. This router or Layer 3 switch is called IGMP querier.

However, a Layer 2 multicast switch does not support IGMP, and therefore cannot send general queries by default. By enabling IGMP Snooping on a Layer 2 switch in a VLAN where multicast traffic needs to be Layer-2 switched only and no multicast

routers are present, the Layer 2 switch will act as the IGMP Snooping querier to send IGMP queries, thus allowing multicast forwarding entries to be established and maintained at the data link layer.

Follow these steps to enable IGMP Snooping querier:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter VLAN view	vlan <i>vlan-id</i>	-
Enable IGMP Snooping querier	igmp-snooping querier	Required Disabled by default



CAUTION: *It is meaningless to configure an IGMP Snooping querier in a multicast network running IGMP. Although an IGMP Snooping querier does not take part in IGMP querier elections, it may affect IGMP querier elections because it sends IGMP general queries with a low source IP address.*

Configuring IGMP Queries and Responses

You can tune the IGMP general query interval based on actual condition of the network.

Upon receiving an IGMP query (general query or group-specific query), a host starts a timer for each multicast group it has joined. This timer is initialized to a random value in the range of 0 to the maximum response time (the host obtains the value of the maximum response time from the Max Response Time field in the IGMP query it received). When the timer value comes down to 0, the host sends an IGMP report to the corresponding multicast group.

An appropriate setting of the maximum response time for IGMP queries allows hosts to respond to queries quickly and avoids bursts of IGMP traffic on the network caused by reports simultaneously sent by a large number of hosts when the corresponding timers expire simultaneously.

- For IGMP general queries, you can configure the maximum response time to fill their Max Response time field.
- For IGMP group-specific queries, you can configure the IGMP last-member query interval to fill their Max Response time field. Namely, for IGMP group-specific queries, the maximum response time equals to the IGMP last-member query interval.

Configuring IGMP queries and responses globally

Follow these steps to configure IGMP queries and responses globally:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter IGMP Snooping view	igmp-snooping	-
Configure the maximum response time to IGMP general queries	max-response-time <i>interval</i>	Optional 10 seconds by default
Configure the IGMP last-member query interval	last-member-query-inter val <i>interval</i>	Optional 1 second by default

Configuring IGMP queries and responses in a VLAN

Follow these steps to configure IGMP queries and responses in a VLAN:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter VLAN view	vlan <i>vlan-id</i>	-
Configure IGMP general query interval	igmp-snooping query-interval <i>interval</i>	Optional 60 seconds by default
Configure the maximum response time to IGMP general queries	igmp-snooping max-response-time <i>interval</i>	Optional 10 seconds by default
Configure the IGMP last-member query interval	igmp-snooping last-member-query-interval <i>interval</i>	Optional 1 second by default



CAUTION: In the configuration, make sure that the IGMP general query interval is larger than the maximum response time for IGMP general queries. Otherwise, multicast group members may be deleted by mistake.

Configuring Source IP Address of IGMP Queries

Upon receiving an IGMP query whose source IP address is 0.0.0.0 on a port, the switch will not set that port as a router port. This may prevent multicast forwarding entries from being correctly created at the data link layer and cause multicast traffic forwarding failure in the end. When a Layer 2 device acts as an IGMP-Snooping querier, to avoid the aforesaid problem, you are commended to configure a non-all-zero IP address as the source IP address of IGMP queries.

Follow these steps to configure source IP address of IGMP queries:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter VLAN view	vlan <i>vlan-id</i>	-
Configure the source address of IGMP general queries	igmp-snooping general-query source-ip { current-interface <i>ip-address</i> }	Optional 0.0.0.0 by default
Configure the source IP address of IGMP group-specific queries	igmp-snooping special-query source-ip { current-interface <i>ip-address</i> }	Optional 0.0.0.0 by default



CAUTION: The source address of IGMP query messages may affect IGMP querier selection within the segment.

Configuring an IGMP Snooping Policy

Configuration Prerequisites

Before configuring an IGMP Snooping policy, complete the following task:

- Enable IGMP Snooping in the VLAN or enable IGMP on the desired VLAN interface

Before configuring an IGMP Snooping policy, prepare the following data:

- ACL rule for multicast group filtering
- The maximum number of multicast groups that can pass the ports

Configuring a Multicast Group Filter

On an IGMP Snooping-enabled switch, the configuration of a multicast group allows the service provider to define restrictions on multicast programs available to different users.

In an actual application, when a user requests a multicast program, the user's host initiates an IGMP report. Upon receiving this report message, the switch checks the report against the configured ACL rule. If the port on which the report was heard can join this multicast group, the switch adds an entry for this port in the IGMP Snooping forwarding table; otherwise the switch drops this report message. Any multicast data that has failed the ACL check will not be sent to this port. In this way, the service provider can control the VOD programs provided for multicast users.

Configuring a multicast group filter globally

Follow these steps to configure a multicast group filter globally:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter IGMP Snooping view	igmp-snooping	-
Configure a multicast group filter	group-policy <i>acl-number</i> [vlan <i>vlan-list</i>]	Required No group filter is configured by default, namely hosts can join any multicast group.

Configuring a multicast group filter on a port or a group of ports

Follow these steps to configuring a multicast group filter on a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter the corresponding view	Enter Ethernet port view interface <i>interface-type</i> <i>interface-number</i>	Use either command
	Enter port group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	
Configure a multicast group filter	igmp-snooping group-policy <i>acl-number</i> [vlan <i>vlan-list</i>]	Required No filter is configured by default, namely hosts can join any multicast group.

Configuring Multicast Source Port Filtering

With the multicast source port filtering feature enabled on a port, the port can be connected with multicast receivers only rather than with multicast sources, because the port will block all multicast data packets while it permits multicast protocol packets to pass.

If this feature is disabled on a port, the port can be connected with both multicast sources and multicast receivers.

Configuring multicast source port filtering globally

Follow these steps to configure multicast source port filtering globally:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter IGMP Snooping view	igmp-snooping	-
Enable multicast source port filtering	source-deny port interface-list	Required Disabled by default

Configuring multicast source port filtering on a port or a group of ports

Follow these steps to configure multicast source port filtering on a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter the corresponding view	Enter Ethernet port view interface <i>interface-type interface-number</i> Enter port group view port-group { manual port-group-name aggregation agg-id }	Use either command
Enable multicast source port filtering	igmp-snooping source-deny	Required Disabled by default



When enabled to filter IPv4 multicast data based on the source ports, the device is automatically enabled to filter IPv6 multicast data based on the source ports.

Configuring the Function of Dropping Unknown Multicast Data

Unknown multicast data refers to multicast data for which no entries exist in the IGMP Snooping forwarding table. When the switch receives such multicast traffic:

- With the function of dropping unknown multicast data enabled, the switch drops all the unknown multicast data received.
- With the function of dropping unknown multicast data disabled, the switch floods unknown multicast data in the VLAN which the unknown multicast data belongs to.

Follow these steps to configure the function of dropping unknown multicast data in a VLAN:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter VLAN view	vlan <i>vlan-id</i>	-
Enable the function of dropping unknown multicast data	igmp-snooping drop-unknown	Required Disabled by default



When enabled to drop unknown IPv4 multicast data, the device is automatically enabled to drop unknown IPv6 multicast data.

Configuring IGMP Report Suppression

When a Layer 2 device receives an IGMP report from a multicast group member, the device forwards the message to the Layer 3 device directly connected with it. Thus, when multiple members of a multicast group are attached to the Layer 2 device, the Layer 3 device directly connected with it will receive duplicate IGMP reports from these members.

With the IGMP report suppression function enabled, within each query cycle, the Layer 2 device forwards only the first IGMP report per multicast group to the Layer 3 device and will not forward the subsequent IGMP reports from the same multicast group to the Layer 3 device. This helps reduce the number of packets being transmitted over the network.

Follow these steps to configure IGMP report suppression:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter IGMP Snooping view	igmp-snooping	-
Enable IGMP report suppression	report-aggregation	Optional Enabled by default

Configuring Maximum Multicast Groups that Can Be Joined on a Port

By configuring the maximum number of multicast groups that can be joined on a port, you can limit the number of multicast programs on-demand available to users, thus to regulate traffic on the port.

Follow these steps to configure the maximum number of multicast groups that can be joined on a port or ports:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter the corresponding view	Enter Ethernet port view interface <i>interface-type interface-number</i>	Use either command
Enter port group view	Enter port group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	
Configure the maximum number of multicast groups that can be joined on the port(s)	igmp-snooping group-limit <i>limit</i> [vlan <i>vlan-list</i>]	Optional The default is 1024.



- *When the number of multicast groups a port has joined reaches the maximum number configured, the system deletes all the forwarding entries persistent to that port from the IGMP Snooping forwarding table, and the hosts on this port need to join the multicast groups again.*
- *If you have configured static or simulated joins on a port, however, when the number of multicast groups on the port exceeds the configured threshold, the system deletes all the forwarding entries persistent to that port from the IGMP Snooping forwarding table and applies the static or simulated joins again, until the number of multicast groups joined by the port comes back within the configured threshold.*

Configuring Multicast Group Replacement

For some special reasons, the number of multicast groups that can be joined on the current switch or port may exceed the number configured for the switch or the port. In addition, in some specific applications, a multicast group newly joined on the switch needs to replace an existing multicast group automatically. A typical example is “channel switching”, namely, by joining a new multicast group, a user automatically switches from the current multicast group to the new one.

To address such situations, you can enable the multicast group replacement function on the switch or certain ports. When the number of multicast groups joined on the switch or a port has joined reaches the limit:

- If the multicast group replacement feature is enabled, the newly joined multicast group automatically replaces an existing multicast group with the lowest address.
- If the multicast group replacement feature is not enabled, new IGMP reports will be automatically discarded.

Configuring multicast group replacement globally

Follow these steps to configure multicast group replacement globally:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter IGMP Snooping view	igmp-snooping	-
Configure multicast group replacement	overflow-replace [vlan <i>vlan-list</i>]	Required Disabled by default

Configuring multicast group replacement on a port or a group of ports

Follow these steps to configure multicast group replacement on a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter the corresponding view	Enter Ethernet port view interface <i>interface-type</i> <i>interface-number</i>	Use either command
	Enter port group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	
Configure multicast group replacement	igmp-snooping overflow-replace [vlan <i>vlan-list</i>]	Required Disabled by default



CAUTION: Be sure to configure the maximum number of multicast groups allowed on a port (refer to “Configuring Maximum Multicast Groups that Can Be Joined on a Port” on page 568) before configuring multicast group replacement. Otherwise, the multicast group replacement functionality will not take effect.

Displaying and Maintaining IGMP Snooping

To do...	Use the command...	Remarks
View the information of IGMP Snooping multicast groups	display igmp-snooping group [vlan <i>vlan-id</i>] [verbose]	Available in any view

To do...	Use the command...	Remarks
View the statistics information of IGMP messages learned by IGMP Snooping	display igmp-snooping statistics	Available in any view
Clear IGMP Snooping multicast group information	reset igmp-snooping group { <i>group-address</i> all } [vlan <i>vlan-id</i>]	Available in user view
Clear the statistics information of all kinds of IGMP messages learned by IGMP Snooping	reset igmp-snooping statistics	Available in user view



- The **reset igmp-snooping group** command works only on an IGMP Snooping-enabled VLAN, but not on a VLAN with IGMP enabled on its VLAN interface.
- The **reset igmp-snooping group** command cannot clear IGMP Snooping forwarding table entries for static joins.

IGMP Snooping Configuration Examples

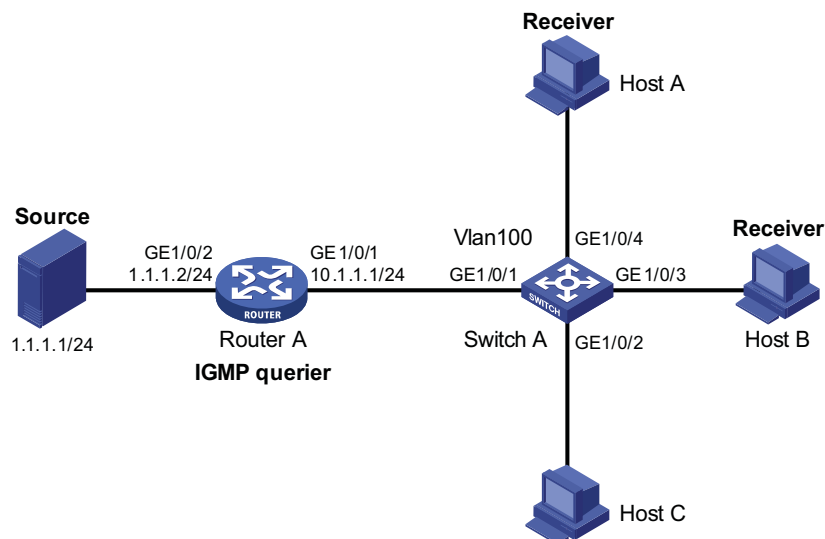
Configuring Simulated Joining

Network requirements

- As shown in Figure 172, Router A connects to the multicast source through GigabitEthernet 1/0/2 and to Switch A through GigabitEthernet 1/0/1. Router A is the IGMP querier on the subnet.
- IGMP is required on Router A, IGMP Snooping is required on Switch A, and Router A will act as the IGMP querier on the subnet.
- Perform the following configuration so that multicast data can be forwarded through GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 even if Host A and Host B temporarily stop receiving multicast data for some unexpected reasons.

Network diagram

Figure 172 Network diagram for simulated joining configuration



Configuration procedure

1 Configure the IP address of each interface

Configure an IP address and subnet mask for each interface as per Figure 172. The detailed configuration steps are omitted.

2 Configure Router A

Enable IP multicast routing, enable PIM-DM on each interface, and enable IGMPv2 on GigabitEthernet 1/0/1.

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface GigabitEthernet 1/0/1
[RouterA-GigabitEthernet1/0/1] igmp enable
[RouterA-GigabitEthernet1/0/1] pim dm
[RouterA-GigabitEthernet1/0/1] quit
[RouterA] interface GigabitEthernet 1/0/2
[RouterA-GigabitEthernet1/0/2] pim dm
[RouterA-GigabitEthernet1/0/2] quit
```

3 Configure Switch A

Enable IGMP Snooping globally.

```
<SwitchA> system-view
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] quit
```

Create VLAN 100, assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to this VLAN, and enable IGMP Snooping in the VLAN.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port GigabitEthernet 1/0/1 to GigabitEthernet 1/0/4
[SwitchA-vlan100] igmp-snooping enable
[SwitchA-vlan100] quit
```

Enable simulated host joining on GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 respectively.

```
[SwitchA] interface GigabitEthernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] igmp-snooping host-join 224.1.1.1 vlan 100
[SwitchA-GigabitEthernet1/0/3] quit
[SwitchA] interface GigabitEthernet 1/0/4
[SwitchA-GigabitEthernet1/0/4] igmp-snooping host-join 224.1.1.1 vlan 100
[SwitchA-GigabitEthernet1/0/4] quit
```

4 Verify the configuration

View the detailed information about IGMP Snooping multicast groups in VLAN 100 on Switch A.

```
[SwitchA] display igmp-snooping group vlan 100 verbose
Total 1 IP Group(s).
Total 1 IP Source(s).
Total 1 MAC Group(s).
```

Port flags: D-Dynamic port, S-Static port, A-Aggregation port, C-Copy port
Subvlan flags: R-Real VLAN, C-Copy VLAN

Vlan(id):100.

```
Total 1 IP Group(s).
Total 1 IP Source(s).
Total 1 MAC Group(s).
```

Router port(s):total 1 port.

```
GE1/0/1 (D) ( 00:01:30 )
```

IP group(s):the following ip group(s) match to one mac group.

```
IP group address:224.1.1.1
```

```

(0.0.0.0, 224.1.1.1):
  Attribute:      Host Port
  Host port(s):total 2 port.
    GE1/0/3              (D) ( 00:03:23 )
    GE1/0/4              (D) ( 00:03:23 )
MAC group(s):
  MAC group address:0100-5e01-0101
  Host port(s):total 2 port.
    GE1/0/3
    GE1/0/4

```

As shown above, GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 of Switch A have joined multicast group 224.1.1.1.

Static Router Port Configuration

Network requirements

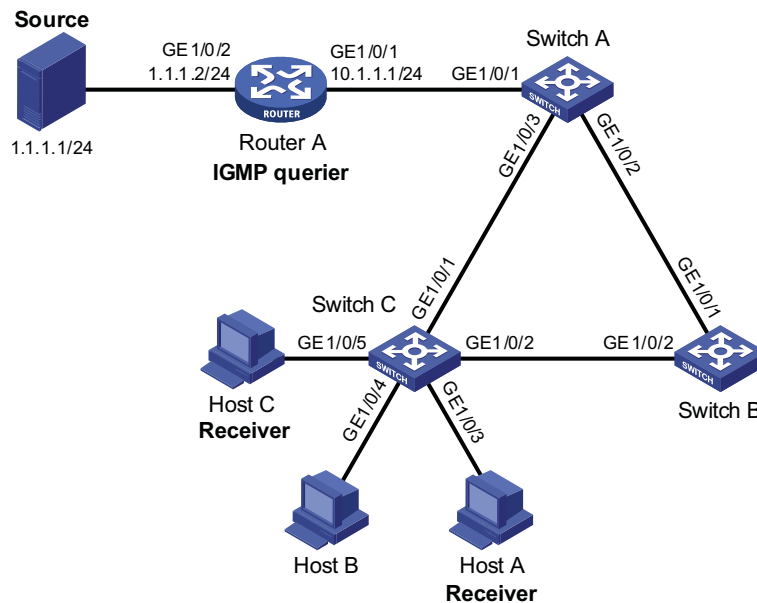
- As shown in Figure 173, Router A connects to a multicast source (Source) through GigabitEthernet 1/0/2, and to Switch A through GigabitEthernet 1/0/1.
- IGMP is to run between Router A and Switch A, and IGMP Snooping is to run on Switch A, Switch B and Switch C, with Router A acting as the IGMP querier.
- Suppose STP runs on the network. To avoid data loops, the forwarding path from Switch A to Switch C is blocked under normal conditions, and multicast traffic flows to the receivers, Host A and Host C, attached to Switch C only along the path of Switch A-Switch B-Switch C.
- Now it is required to configure GigabitEthernet 1/0/3 that connects Switch A to Switch C as a static router port, so that multicast traffic can flow to the receivers nearly uninterruptedly along the path of Switch A-Switch C in the case that the path of Switch A-Switch B-Switch C gets blocked.



If no static router port is configured, when the path of Switch A-Switch B-Switch C gets blocked, at least one IGMP query-response cycle must be completed before the multicast data can flow to the receivers along the new path of Switch A-Switch C, namely multicast delivery will be interrupted during this process.

Network diagram

Figure 173 Network diagram for static router port configuration



Configuration procedure

- 1 Configure the IP address of each interface

Configure an IP address and subnet mask for each interface as per Figure 173. The detailed configuration steps are omitted.

- 2 Configure Router A

Enable IP multicast routing, enable PIM-DM on each interface, and enable IGMP on GigabitEthernet 1/0/1.

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface GigabitEthernet 1/0/1
[RouterA-GigabitEthernet1/0/1] igmp enable
[RouterA-GigabitEthernet1/0/1] pim dm
[RouterA-GigabitEthernet1/0/1] quit
[RouterA] interface GigabitEthernet 1/0/2
[RouterA-GigabitEthernet1/0/2] pim dm
[RouterA-GigabitEthernet1/0/2] quit
```

- 3 Configure Switch A

Enable IGMP Snooping globally.

```
<SwitchA> system-view
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] quit
```

Create VLAN 100, assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to this VLAN, and enable IGMP Snooping in the VLAN.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port GigabitEthernet 1/0/1 to GigabitEthernet 1/0/3
[SwitchA-vlan100] igmp-snooping enable
[SwitchA-vlan100] quit
```

Configure GigabitEthernet 1/0/3 to be a static router port.

```
[SwitchA] interface GigabitEthernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] igmp-snooping static-router-port vlan
100
[SwitchA-GigabitEthernet1/0/3] quit
```

4 Configure Switch B

Enable IGMP Snooping globally.

```
<SwitchB> system-view
[SwitchB] igmp-snooping
[SwitchB-igmp-snooping] quit
```

Create VLAN 100, assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to this VLAN, and enable IGMP Snooping in the VLAN.

```
[SwitchB] vlan 100
[SwitchB-vlan100] port GigabitEthernet 1/0/1 GigabitEthernet 1/0/2
[SwitchB-vlan100] igmp-snooping enable
[SwitchB-vlan100] quit
```

5 Configure Switch C

Enable IGMP Snooping globally.

```
<SwitchC> system-view
[SwitchC] igmp-snooping
[SwitchC-igmp-snooping] quit
```

Create VLAN 100, assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/5 to this VLAN, and enable IGMP Snooping in the VLAN.

```
[SwitchC] vlan 100
[SwitchC-vlan100] port GigabitEthernet 1/0/1 to GigabitEthernet 1/0/
5
[SwitchC-vlan100] igmp-snooping enable
[SwitchC-vlan100] quit
```

6 Verify the configuration

View the detailed information about IGMP Snooping multicast groups in VLAN 100 on Switch A.

```
[SwitchA] display igmp-snooping group vlan 100 verbose
Total 1 IP Group(s).
Total 1 IP Source(s).
Total 1 MAC Group(s).
```

Port flags: D-Dynamic port, S-Static port, A-Aggregation port, C-Copy port

Subvlan flags: R-Real VLAN, C-Copy VLAN

Vlan(id):100.

Total 1 IP Group(s).

Total 1 IP Source(s).

Total 1 MAC Group(s).

Router port(s):total 2 port.

GE1/0/1 (D) (00:01:30)

GE1/0/3 (S)

IP group(s):the following ip group(s) match to one mac group.

IP group address:224.1.1.1

(0.0.0.0, 224.1.1.1):

Attribute: Host Port

```

Host port(s):total 1 port.
  GE1/0/2                (D) ( 00:03:23 )
MAC group(s):
  MAC group address:0100-5e01-0101
  Host port(s):total 1 port.
  GE1/0/2

```

As shown above, GigabitEthernet 1/0/3 of Switch A has become a static router port.

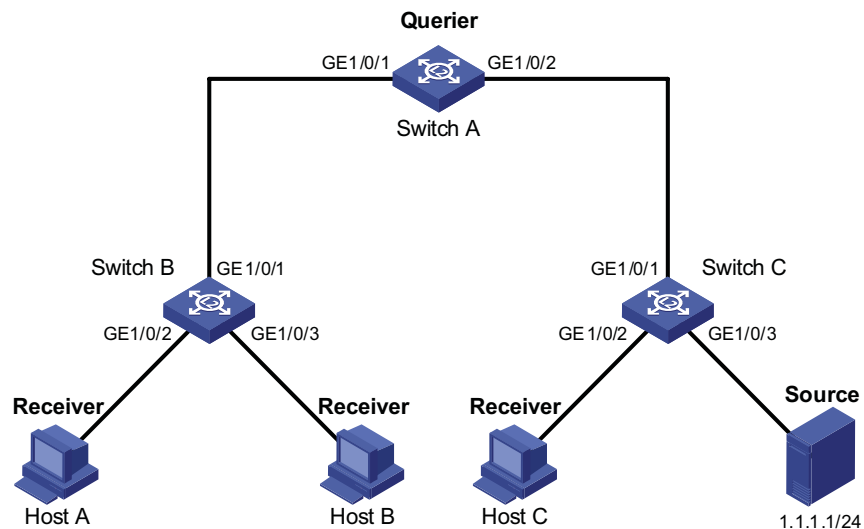
IGMP Snooping Querier Configuration

Network requirements

- As shown in Figure 174, in a Layer-2-only network environment, Switch C is connected to the multicast source (Source) through GigabitEthernet 1/0/3. At least one receiver is attached to Switch B and Switch C respectively.
- IGMPv2 is enabled on all the receivers. Switch A, Switch B, and Switch C run IGMP Snooping. Switch A acts as the IGMP-Snooping querier.
- Configure a non-all-zero IP address as the source IP address of IGMP queries to ensure normal creation of multicast forwarding entries.

Network diagram

Figure 174 Network diagram for IGMP Snooping querier configuration



Configuration procedure

1 Configure switch A

Enable IGMP Snooping globally.

```

<SwitchA> system-view
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] quit

```

Create VLAN 100 and add GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to VLAN 100.

```

[SwitchA] vlan 100
[SwitchA-vlan100] port GigabitEthernet 1/0/1 GigabitEthernet 1/0/2

```

Enable IGMP Snooping in VLAN 100 and configure the IGMP-Snooping querier feature.

```
[SwitchA-vlan100] igmp-snooping enable
[SwitchA-vlan100] igmp-snooping querier

# Set the source IP address of IGMP general queries and group-specific queries to
192.168.1.1.

[SwitchA-vlan100] igmp-snooping general-query source-ip 192.168.1.1
[SwitchA-vlan100] igmp-snooping special-query source-ip 192.168.1.1
```

2 Configure Switch B

Enable IGMP Snooping globally.

```
<SwitchB> system-view
[SwitchB] igmp-snooping
[SwitchB-igmp-snooping] quit
```

Create VLAN 100, add GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to VLAN 100, and enable IGMP Snooping in this VLAN.

```
[SwitchB] vlan 100
[SwitchB-vlan100] port GigabitEthernet 1/0/1 to GigabitEthernet 1/0/3
[SwitchB-vlan100] igmp-snooping enable
```

3 Configuration on Switch C

Enable IGMP Snooping globally.

```
<SwitchC> system-view
[SwitchC] igmp-snooping
[SwitchC-igmp-snooping] quit
```

Create VLAN 100, add GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to VLAN 100, and enable IGMP Snooping in this VLAN.

```
[SwitchC] vlan 100
[SwitchC-vlan100] port GigabitEthernet 1/0/1 to GigabitEthernet 1/0/3
[SwitchC-vlan100] igmp-snooping enable
```

4 Verify the configuration

View the IGMP message statistics on Switch C.

```
[SwitchC-vlan100] display igmp-snooping statistics
Received IGMP general queries:3.
Received IGMPv1 reports:0.
Received IGMPv2 reports:4.
Received IGMP leaves:0.
Received IGMPv2 specific queries:0.
Sent IGMPv2 specific queries:0.
Received IGMPv3 reports:0.
Received IGMPv3 reports with right and wrong records:0.
Received IGMPv3 specific queries:0.
Received IGMPv3 specific sg queries:0.
Sent IGMPv3 specific queries:0.
Sent IGMPv3 specific sg queries:0.
Received error IGMP messages:0.
```

Switch C received IGMP general queries. This means that Switch A works as an IGMP-Snooping querier.

Troubleshooting IGMP Snooping Configuration

Switch Fails in Layer 2 Multicast Forwarding

Symptom

A switch fails to implement Layer 2 multicast forwarding.

Analysis

IGMP Snooping is not enabled.

Solution

- 1 Enter the **display current-configuration** command to view the running status of IGMP Snooping.
- 2 If IGMP Snooping is not enabled, use the **igmp-snooping** command to enable IGMP Snooping globally, and then use **igmp-snooping enable** command to enable IGMP Snooping in VLAN view.
- 3 If IGMP Snooping is disabled only for the corresponding VLAN, just use the **igmp-snooping enable** command in VLAN view to enable IGMP Snooping in the corresponding VLAN.

Configured Multicast Group Policy Fails to Take Effect

Symptom

Although a multicast group policy has been configured to allow hosts to join specific multicast groups, the hosts can still receive multicast data addressed to other multicast groups.

Analysis

- The ACL rule is incorrectly configured.
- The multicast group policy is not correctly applied.
- The function of dropping unknown multicast data is not enabled, so unknown multicast data is flooded.
- Certain ports have been configured as static member ports of multicasts groups, and this configuration conflicts with the configured multicast group policy.

Solution

- 1 Use the **display acl** command to check the configured ACL rule. Make sure that the ACL rule conforms to the multicast group policy to be implemented.
- 2 Use the **display this** command in IGMP Snooping view or in the corresponding interface view to check whether the correct multicast group policy has been applied. If not, use the **group-policy** or **igmp-snooping group-policy** command to apply the correct multicast group policy.
- 3 Use the **display current-configuration** command to check whether the function of dropping unknown multicast data is enabled. If not, use the **igmp-snooping drop-unknown** command to enable the function of dropping unknown multicast data.
- 4 Use the **display igmp-snooping group** command to check whether any port has been configured as a static member port of any multicast group. If so, check

whether this configuration conflicts with the configured multicast group policy. If any conflict exists, remove the port as a static member of the multicast group.

43

MLD SNOOPING CONFIGURATION

When configuring MLD Snooping, go to these sections for information you are interested in:

- “MLD Snooping Overview” on page 579
- “MLD Snooping Configuration Task List” on page 583
- “Displaying and Maintaining MLD Snooping” on page 595
- “MLD Snooping Configuration Examples” on page 596
- “Troubleshooting MLD Snooping” on page 602

MLD Snooping Overview

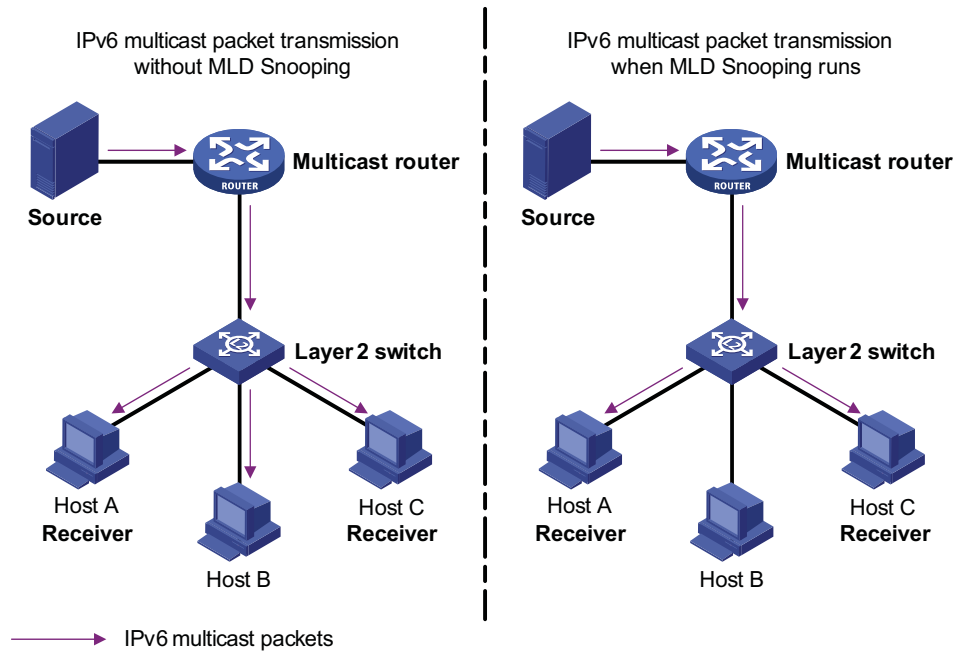
Multicast Listener Discovery Snooping (MLD Snooping) is an IPv6 multicast constraining mechanism that runs on Layer 2 devices to manage and control IPv6 multicast groups.

Introduction to MLD Snooping

By analyzing received MLD messages, a Layer 2 device running MLD Snooping establishes mappings between ports and multicast MAC addresses and forwards IPv6 multicast data based on these mappings.

As shown in Figure 175, when MLD Snooping is not running, IPv6 multicast packets are broadcast to all devices at Layer 2. When MLD Snooping runs, multicast packets for known IPv6 multicast groups are multicast to the receivers at Layer 2.

Figure 175 Before and after MLD Snooping is enabled on the Layer 2 device

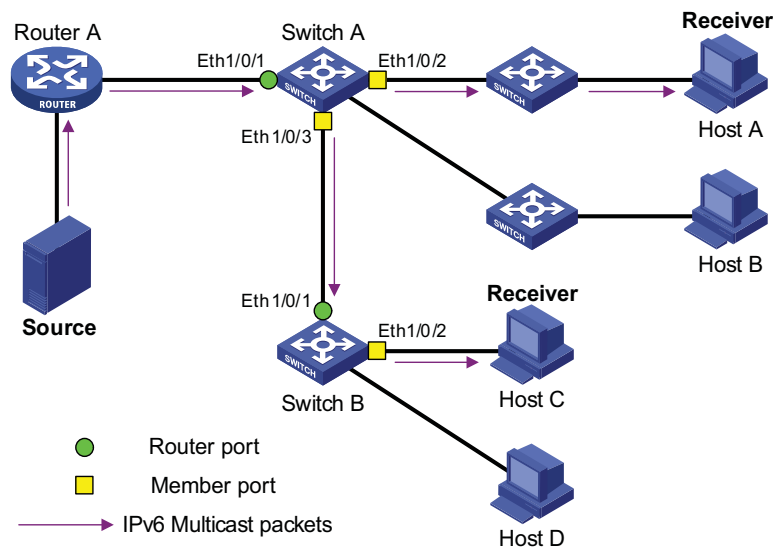


Basic Concepts in MLD Snooping

MLD Snooping related ports

As shown in Figure 171, Router A connects to the multicast source, MLD Snooping runs on Switch A and Switch B, Host A and Host C are receiver hosts (namely, IPv6 multicast group members).

Figure 176 MLD Snooping related ports



Ports involved in MLD Snooping, as shown in Figure 171, are described as follows:

- Router port: A router port is a port on the Ethernet switch that leads switch towards the Layer-3 multicast device (DR or MLD querier). In the figure, Ethernet 1/0/1 of Switch A and Ethernet 1/0/1 of Switch B are router ports. The

switch registers all its local router ports (including static and dynamic router ports) in its router port list.

- Member port: A member port (also known as IPv6 multicast group member port) is a port on the Ethernet switch that leads switch towards multicast group members. In the figure, Ethernet 1/0/2 and Ethernet 1/0/3 of Switch A and Ethernet 1/0/2 of Switch B are member ports. The switch registers all the member ports (including static and dynamic member ports) on the local device in its MLD Snooping forwarding table.



- *Whenever mentioned in this document, a router port is a router-connecting port on the switch, rather than a port on a router.*
- *On an MLD-snooping-enabled switch, the ports that received MLD general queries with the source address other than 0::0 or IPv6 PIM hello messages are router ports.*

Aging timers for dynamic ports in MLD Snooping

Table 56 Aging timers for dynamic ports in MLD Snooping and related messages and actions

Timer	Description	Message before expiry	Action after expiry
Router port aging timer	For each router port, the switch sets a timer initialized to the aging time of the route port.	MLD general query of which the source address is not 0::0 or IPv6 PIM hello.	The switch removes this port from its router port list.
Member port aging timer	When a port joins an IPv6 multicast group, the switch sets a timer for the port, which is initialized to the member port aging time.	MLD report message.	The switch removes this port from the IPv6 multicast group forwarding table.



The port aging mechanism of MLD Snooping works only for dynamic ports; a static port will never age out.

How MLD Snooping Works

A switch running MLD Snooping performs different actions when it receives different MLD messages, as follows:

General queries

The MLD querier periodically sends MLD general queries to all hosts and routers (FF02::1) on the local subnet to find out whether IPv6 multicast group members exist on the subnet.

Upon receiving an MLD general query, the switch forwards it through all ports in the VLAN except the receiving port and performs the following to the receiving port:

- If the receiving port is a router port existing in its router port list, the switch resets the aging timer of this router port.
- If the receiving port is not a router port existing in its router port list, the switch adds it into its router port list and sets an aging timer for this router port.

Membership reports

A host sends an MLD report to the multicast router in the following circumstances:

- Upon receiving an MLD query, an IPv6 multicast group member host responds with an MLD report.
- When intended to join an IPv6 multicast group, a host sends an MLD report to the multicast router to announce that it is interested in the multicast information addressed to that IPv6 multicast group.

Upon receiving an MLD report, the switch forwards it through all the router ports in the VLAN, resolves the address of the reported IPv6 multicast group, and performs the following to the receiving port:

- If no forwarding table entry exists for the reported IPv6 multicast group, the switch creates an entry, adds the port as member port to the outgoing port list, and starts a member port aging timer for that port.
- If a forwarding table entry exists for the reported IPv6 multicast group, but the port is not included in the outgoing port list for that group, the switch adds the port as a member port to the outgoing port list, and starts a member port aging timer for that port.
- If a forwarding table entry exists for the reported IPv6 multicast group and the port is included in the outgoing port list, which means that this port is already a member port, the switch resets the member port aging timer for that port.



A switch does not forward an MLD report through a non-router port. The reason is as follows: Due to the MLD report suppression mechanism, if the switch forwards a report message through a member port, all the attached hosts listening to the reported IPv6 multicast address will suppress their own reports upon hearing this report, and this will prevent the switch from knowing whether any hosts attached to that port are still active members of the reported IPv6 multicast group.

Done messages

When a host leaves an IPv6 multicast group, the host sends an MLD done message to the multicast router.

When the switch receives a group-specific MLD done message on a member port, it first checks whether a forwarding table entry for that IPv6 multicast group exists, and, if one exists, whether its outgoing port list contains that port.

- If the forwarding table entry does not exist or if its outgoing port list does not contain the port, the switch discards the MLD done message instead of forwarding it to any port.
- If the forwarding table entry exists and its outgoing port list contains the port, the switch forwards the done message to all router ports in the VLAN. Because the switch does not know whether any other hosts attached to the port are still listening to that IPv6 multicast group address, the switch does not immediately remove the port from the outgoing port list of the forwarding table entry for that group; instead, it resets the member port aging timer for the port.

Upon receiving an MLD done message from a host, the MLD querier resolves from the message the address of the IPv6 multicast group that the host just left and

sends an MLD multicast-address-specific query to that IPv6 multicast group through the port that received the done message. Upon hearing the MLD multicast-address-specific query, the switch forwards it through all its router ports in the VLAN and all member ports for that IPv6 multicast group, and performs the following to the receiving port:

- If any MLD report in response to the MLD multicast-address-specific query is heard on a member port before its aging timer expires, this means that some host attached to the port is receiving or expecting to receive IPv6 multicast data for that IPv6 multicast group. The switch resets the aging timer of the member port.
- If no MLD report in response to the MLD multicast-address-specific query is heard on a member port before its aging timer expires, this means that no hosts attached to the port are still listening to that IPv6 multicast group address. The switch removes the port from the outgoing port list of the forwarding table entry for that IPv6 multicast group when the aging timer expires.

Protocols and Standards MLD Snooping is documented in:

- RFC 4541: Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches

MLD Snooping Configuration Task List

Complete these tasks to configure MLD Snooping:

Task	Remarks
"Configuring Basic Functions of MLD Snooping" on page 584	"Enabling MLD Snooping" on page 584 Required
	"Configuring the Version of MLD Snooping" on page 585 Optional
"Configuring MLD Snooping Port Functions" on page 585	"Configuring Aging Timers for Dynamic Ports" on page 586 Optional
	"Configuring Static Ports" on page 586 Optional
	"Configuring Simulated Joining" on page 587 Optional
	"Configuring Fast Leave Processing" on page 588 Optional
"Configuring MLD Snooping Querier" on page 589	"Enabling MLD Snooping Querier" on page 589 Optional
	"Configuring MLD Queries and Responses" on page 590 Optional
	"Configuring Source IPv6 Addresses of MLD Queries" on page 591 Optional

Task	Remarks
"Configuring an MLD Snooping Policy" on page 591	Optional
"Configuring an IPv6 Multicast Group Filter" on page 591	Optional
"Configuring IPv6 Multicast Source Port Filtering" on page 592	Optional
"Configuring Dropping Unknown IPv6 Multicast Data" on page 593	Optional
"Configuring MLD Report Suppression" on page 593	Optional
"Configuring Maximum Multicast Groups that Can Be Joined on a Port" on page 594	Optional
"Configuring IPv6 Multicast Group Replacement" on page 594	Optional



- Configurations made in MLD Snooping view are effective for all VLANs, while configurations made in VLAN view are effective only for ports belonging to the current VLAN. For a given VLAN, a configuration made in MLD Snooping view is effective only if the same configuration is not made in VLAN view.
- Configurations made in MLD Snooping view are effective for all ports; configurations made in Ethernet port view are effective only for the current port; configurations made in manual port group view are effective only for all the ports in the current port group; configurations made in aggregation group view are effective only for the master port. For a given port, a configuration made in MLD Snooping view is effective only if the same configuration is not made in Ethernet port view or port group view.

Configuring Basic Functions of MLD Snooping

Configuration Prerequisites

Before configuring the basic functions of MLD Snooping, complete the following tasks:

- Configure the corresponding VLANs

Before configuring the basic functions of MLD Snooping, prepare the following data:

- The version of MLD Snooping

Enabling MLD Snooping

Follow these steps to enable MLD Snooping:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable MLD Snooping globally and enter MLD-Snooping view	mld-snooping	Required Disabled by default
Return to system view	quit	-

To do...	Use the command...	Remarks
Enter VLAN view	vlan <i>vlan-id</i>	-
Enable MLD Snooping in the VLAN	mld-snooping enable	Required Disabled by default



- *MLD Snooping must be enabled globally before it can be enabled in a VLAN.*
- *After enabling MLD Snooping in a VLAN, you cannot enable MLD and/or IPv6 PIM on the corresponding VLAN interface, and vice versa.*
- *When you enable MLD Snooping in a specified VLAN, this function takes effect for Ethernet ports in this VLAN only.*

Configuring the Version of MLD Snooping

By configuring the MLD Snooping version, you actually configure the version of MLD messages that MLD Snooping can process.

- MLD Snooping version 1 can process MLDv1 messages, but cannot analyze and process MLDv2 messages, which will be flooded in the VLAN.
- MLD Snooping version 2 can process MLDv1 and MLDv2 messages.

Follow these steps to configure the version of MLD Snooping:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter VLAN view	vlan <i>vlan-id</i>	-
Configure the version of MLD Snooping	mld-snooping version <i>version-number</i>	Optional Version 1 by default



CAUTION: *If you switch MLD Snooping from version 2 to version 1, the system will clear all MLD Snooping forwarding entries from dynamic joins, and will*

- *Keep forwarding entries from version 2 static (*, G) joins;*
- *Clear forwarding entries from version 2 static (S, G) joins, which will be restored when MLD Snooping is switched back to version 2.*

For details about static joins, Refer to “Configuring Static Ports” on page 586.

Configuring MLD Snooping Port Functions

Configuration Prerequisites

Before configuring MLD Snooping port functions, complete the following tasks:

- Enable MLD Snooping in the VLAN
- Configure the corresponding port groups

Before configuring MLD Snooping port functions, prepare the following data:

- Aging time of router ports
- Aging timer of member ports

- IPv6 multicast group and IPv6 multicast source addresses

Configuring Aging Timers for Dynamic Ports

If the switch receives no MLD general queries or IPv6 PIM hello messages on a dynamic router port, the switch removes the port from the router port list when the aging timer of the port expires.

If the switch receives no MLD reports for an IPv6 multicast group on a dynamic member port, the switch removes the port from the outgoing port list of the forwarding table entry for that IPv6 multicast group when the aging timer of the port for that group expires.

If IPv6 multicast group memberships change frequently, you can set a relatively small value for the member port aging timer, and vice versa.

Configuring aging timers for dynamic ports globally

Follow these steps to configure aging timers for dynamic ports globally:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter MLD Snooping view	mld-snooping	-
Configure router port aging time	router-aging-time <i>interval</i>	Optional 260 seconds by default
Configure member port aging time	host-aging-time <i>interval</i>	Optional 260 seconds by default

Configuring aging timers for dynamic ports in a VLAN

Follow these steps to configure aging timers for dynamic ports in a VLAN:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter VLAN view	vlan <i>vlan-id</i>	-
Configure router port aging time	mld-snooping router-aging-time <i>interval</i>	Optional 260 seconds by default
Configure member port aging time	mld-snooping host-aging-time <i>interval</i>	Optional 260 seconds by default

Configuring Static Ports

If all the hosts attached to a port is interested in the IPv6 multicast data addressed to a particular IPv6 multicast group, you can configure that port as a static member port for that IPv6 multicast group.

You can configure a port of a switch to be static router port, through which the switch can forward all IPv6 multicast data it received.

Follow these steps to configure static ports:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...		Use the command...	Remarks
Enter the corresponding view	Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	Use either command
	Enter port group view	port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	
Configure the port(s) as static member port(s)		mld-snooping static-group <i>ipv6-group-address</i> [source-ip <i>ipv6-source-address</i>] vlan <i>vlan-id</i>	Required Disabled by default
Configure the port(s) as static router port(s)		mld-snooping static-router-port <i>vlan vlan-id</i>	Required Disabled by default



- The IPv6 static (S, G) joining function is available only if a valid IPv6 multicast source address is specified and MLD Snooping version 2 is currently running on the switch.
- A static member port does not respond to queries from the MLD querier; when static (*, G) or (S, G) joining is enabled or disabled on a port, the port does not send an unsolicited MLD report or an MLD done message.
- Static member ports and static router ports never age out. To remove such a port, you need to use the corresponding command.

Configuring Simulated Joining

Generally, a host running MLD responds to MLD queries from the MLD querier. If a host fails to respond due to some reasons, the multicast router will deem that no member of this IPv6 multicast group exists on the network segment, and therefore will remove the corresponding forwarding path.

To avoid this situation from happening, you can enable simulated joining on a port of the switch, namely configure the port as a simulated member host for an IPv6 multicast group. When an MLD query is heard, simulated host gives a response. Thus, the switch can continue receiving IPv6 multicast data.

A simulated host acts like a real host, as follows:

- When a port is configured as a simulated member host, the switch sends an unsolicited MLD report through that port.
- After a port is configured as a simulated member host, the switch responds to MLD general queries by sending MLD reports through that port.
- When the simulated joining function is disabled on a port, the switch sends an MLD done message through that port.

Follow these steps to configure simulated joining:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...		Use the command...	Remarks
Enter the corresponding view	Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	Use either command
	Enter port group view	port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	
Configure simulated joining		mld-snooping host-join <i>ipv6-group-address</i> [source-ip <i>ipv6-source-address</i>] vlan <i>vlan-id</i>	Required Disabled by default



- Each simulated host is equivalent to an independent host. For example, when receiving an MLD query, the simulated host corresponding to each configuration responds respectively.
- Unlike a static member port, a port configured as a simulated member host will age out like a dynamic member port.

Configuring Fast Leave Processing

The fast leave processing feature allows the switch to process MLD done messages in a fast way. With the fast leave processing feature enabled, when receiving an MLD done message on a port, the switch immediately removes that port from the outgoing port list of the forwarding table entry for the indicated IPv6 multicast group. Then, when receiving MLD done multicast-address-specific queries for that IPv6 multicast group, the switch will not forward them to that port.

In VLANs where only one host is attached to each port, fast leave processing helps improve bandwidth and resource usage.

Configuring fast leave processing globally

Follow these steps to configure fast leave processing globally:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Enter MLD Snooping view		mld-snooping	-
Enable fast leave processing		fast-leave [vlan <i>vlan-list</i>]	Required Disabled by default

Configuring fast leave processing on a port or a group of ports

Follow these steps to configure fast leave processing on a port or a group of ports:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Enter the corresponding view	Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	Use either command
	Enter port group view	port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	

To do...	Use the command...	Remarks
Enable fast leave processing	mld-snooping fast-leave [vlan vlan-list]	Required Disabled by default



CAUTION: If fast leave processing is enabled on a port to which more than one host is connected, when one host leaves an IPv6 multicast group, the other hosts connected to port and interested in the same IPv6 multicast group will fail to receive IPv6 multicast data addressed to that group.

Configuring MLD Snooping Querier

Configuration Prerequisites

Before configuring MLD Snooping querier, complete the following task:

- Enable MLD Snooping in the VLAN.

Before configuring MLD Snooping querier, prepare the following data:

- MLD general query interval,
- MLD last-member query interval,
- Maximum response time for MLD general queries,
- Source IPv6 address of MLD general queries, and
- Source IPv6 address of MLD multicast-address-specific queries.

Enabling MLD Snooping Querier

In an IPv6 multicast network running MLD, a multicast router or Layer 3 multicast switch is responsible for sending periodic MLD general queries, so that all Layer 3 multicast devices can establish and maintain multicast forwarding entries, thus to forward multicast traffic correctly at the network layer. This router or Layer 3 switch is called MLD querier.

However, a Layer 2 multicast switch does not support MLD, and therefore cannot send MLD general queries by default. By enabling MLD Snooping querier on a Layer 2 switch in a VLAN where multicast traffic needs to be Layer-2 switched only and no Layer 3 multicast devices are present, the Layer 2 switch will act as the MLD querier to send periodic MLD queries, thus allowing multicast forwarding entries to be established and maintained at the data link layer.

Follow these steps to enable the MLD Snooping querier:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter VLAN view	vlan vlan-id	-
Enable the MLD Snooping querier	mld-snooping querier	Required Disabled by default



CAUTION: It is meaningless to configure an MLD Snooping querier in an IPv6 multicast network running MLD. Although an MLD Snooping querier does not

take part in MLD querier elections, it may affect MLD querier elections because it sends MLD general queries with a low source IPv6 address.

Configuring MLD Queries and Responses

You can tune the MLD general query interval based on actual condition of the network.

Upon receiving an MLD query (general query or group-specific query), a host starts a timer for each IPv6 multicast group it has joined. This timer is initialized to a random value in the range of 0 to the maximum response time (the host obtains the value of the maximum response time from the Max Response Time field in the MLD query it received). When the timer value comes down to 0, the host sends an MLD report to the corresponding IPv6 multicast group.

An appropriate setting of the maximum response time for MLD queries allows hosts to respond to queries quickly and avoids burstiness of MLD traffic on the network caused by reports simultaneously sent by a large number of hosts when the corresponding timers expire simultaneously.

- For MLD general queries, you can configure the maximum response time to fill their Max Response time field.
- For MLD multicast-address-specific queries, you can configure the MLD last-member query interval to fill their Max Response time field. Namely, for MLD multicast-address-specific queries, the maximum response time equals to the MLD last-member query interval.

Configuring MLD queries and responses globally

Follow these steps to configure MLD queries and responses globally:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter MLD Snooping view	mld-snooping	-
Configure the maximum response time for MLD general queries	max-response-time <i>interval</i>	Optional 10 seconds by default
Configure the MLD last-member query interval	last-listener-query-interval <i>interval</i>	Optional 1 second by default

Configuring MLD queries and responses in a VLAN

Follow these steps to configure MLD queries and responses in a VLAN

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter VLAN view	vlan <i>vlan-id</i>	-
Configure MLD query interval	mld-snooping query-interval <i>interval</i>	Optional 125 seconds by default
Configure the maximum response time for MLD general queries	mld-snooping max-response-time <i>interval</i>	Optional 10 seconds by default
Configure the MLD last-member query interval	mld-snooping last-listener-query-interval <i>interval</i>	Optional 1 second by default



CAUTION: Make sure that the MLD query interval is greater than the maximum response time for MLD general queries; otherwise undesired deletion of IPv6 multicast members may occur.

Configuring Source IPv6 Addresses of MLD Queries

This configuration allows you to change the source IPv6 address of MLD queries.

Follow these steps to configure source IPv6 addresses of MLD queries:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter VLAN view	vlan <i>vlan-id</i>	-
Configure the source IPv6 address of MLD general queries	mld-snooping general-query source-ip { current-interface <i>ipv6-address</i> }	Optional FE80::02FF:FFFF:FE00:0001 by default
Configure the source IPv6 address of MLD multicast-address-specific queries	mld-snooping special-query source-ip { current-interface <i>ipv6-address</i> }	Optional FE80::02FF:FFFF:FE00:0001 by default



CAUTION: The source IPv6 address of MLD query messages may affect MLD querier election within the segment.

Configuring an MLD Snooping Policy

Configuration Prerequisites

Before configuring an MLD Snooping policy, complete the following tasks:

- Enable MLD Snooping in the VLAN

Before configuring an MLD Snooping policy, prepare the following data:

- IPv6 ACL rule for IPv6 multicast group filtering
- The maximum number of IPv6 multicast groups that can pass the ports

Configuring an IPv6 Multicast Group Filter

On a MLD Snooping-enabled switch, the configuration of an IPv6 multicast group filter allows the service provider to define limits of multicast programs available to different users.

In an actual application, when a user requests a multicast program, the user's host initiates an MLD report. Upon receiving this report message, the switch checks the report against the configured ACL rule. If the port on which the report was heard can join this IPv6 multicast group, the switch adds an entry for this port in the MLD Snooping forwarding table; otherwise the switch drops this report message. Any IPv6 multicast data that fails the ACL check will not be sent to this port. In this way, the service provider can control the VOD programs provided for multicast users.

Configuring an IPv6 multicast group filter globally

Follow these steps to configure an IPv6 multicast group globally:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter MLD Snooping view	mld-snooping	-
Configure an IPv6 multicast group filter	group-policy <i>acl6-number</i> [vlan <i>vlan-list</i>]	Required No IPv6 filter configured by default, namely hosts can join any IPv6 multicast group.

Configuring an IPv6 multicast group filter on a port or a group of ports

Follow these steps to configure an IPv6 multicast group filter on a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter the corresponding view	Enter Ethernet port view interface <i>interface-type</i> <i>interface-number</i>	Use either command
	Enter port group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	
Configure an IPv6 multicast group filter	mld-snooping group-policy <i>acl6-number</i> [vlan <i>vlan-list</i>]	Required No IPv6 filter configured by default, namely hosts can join any IPv6 multicast group.

Configuring IPv6 Multicast Source Port Filtering

With the IPv6 multicast source port filtering feature enabled on a port, the port can be connected with IPv6 multicast receivers only rather than with multicast sources, because the port will block all IPv6 multicast data packets while it permits multicast protocol packets to pass.

If this feature is disabled on a port, the port can be connected with both multicast sources and IPv6 multicast receivers.

Configuring IPv6 multicast source port filtering globally

Follow these steps to configure IPv6 multicast source port filtering:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter MLD Snooping view	mld-snooping	-
Enable IPv6 multicast source port filtering	source-deny port <i>interface-list</i>	Required Disabled by default

Configuring IPv6 multicast source port filtering on a port or a group of ports

Follow these steps to configure IPv6 multicast source port filtering on a port or a group of ports:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Enter the corresponding view	Enter Ethernet port view Enter port group view	interface <i>interface-type</i> <i>interface-number</i> port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Use either command
Enable IPv6 multicast source port filtering		mld-snooping source-deny	Required Disabled by default



When enabled to filter IPv6 multicast data based on the source ports, the device is automatically enabled to filter IPv4 multicast data based on the source ports.

Configuring Dropping Unknown IPv6 Multicast Data

Unknown IPv6 multicast data refers to IPv6 multicast data for which no forwarding entries exist in the MLD Snooping forwarding table: When the switch receives such IPv6 multicast traffic:

- With the function of dropping unknown IPv6 multicast data enabled, the switch drops all unknown IPv6 multicast data received.
- With the function of dropping unknown IPv6 multicast data disabled, the switch floods unknown IPv6 multicast data in the VLAN to which the unknown IPv6 multicast data belongs.

Follow these steps to enable dropping unknown IPv6 multicast data in a VLAN:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Enter VLAN view		vlan <i>vlan-id</i>	-
Enable dropping unknown IPv6 multicast data		mld-snooping drop-unknown	Required Disabled by default



When enabled to drop unknown IPv6 multicast data, the device is automatically enabled to drop unknown IPv4 multicast data.

Configuring MLD Report Suppression

When a Layer 2 device receives an MLD report from an IPv6 multicast group member, the Layer 2 device forwards the message to the Layer 3 device directly connected with it. Thus, when multiple members belonging to an IPv6 multicast group exist on the Layer 2 device, the Layer 3 device directly connected with it will receive duplicate MLD reports from these members.

With the MLD report suppression function enabled, within a query interval, the Layer 2 device forwards only the first MLD report of an IPv6 group to the Layer 3 device and will not forward the subsequent MLD reports from the same multicast group to the Layer 3 device. This helps reduce the number of packets being transmitted over the network.

Follow these steps to configure MLD report suppression:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter MLD Snooping view	mld-snooping	-
Enable MLD report suppression	report-aggregation	Optional Enabled by default

Configuring Maximum Multicast Groups that Can Be Joined on a Port

By configuring the maximum number of IPv6 multicast groups that can be joined on a port or a group of ports, you can limit the number of multicast programs available to VOD users, thus to control the traffic on the port.

Follow these steps configure the maximum number of IPv6 multicast groups that can be joined on a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter the corresponding view	Enter Ethernet port view interface <i>interface-type</i> <i>interface-number</i>	Use either command
	Enter port group view port-group { <i>manual</i> <i>port-group-name</i> aggregation <i>agg-id</i> }	
Configure the maximum number of IPv6 multicast groups that can be joined on a port	mld-snooping group-limit <i>limit</i> [vlan <i>vlan-list</i>]	Optional The default is 512.



- *When the number of IPv6 multicast groups that can be joined on a port reaches the maximum number configured, the system deletes all the forwarding entries persistent to that port from the MLD Snooping forwarding table, and the hosts on this port need to join IPv6 multicast groups again.*
- *If you have configured static or simulated joins on a port, however, when the number of IPv6 multicast groups on the port exceeds the configured threshold, the system deletes all the forwarding entries persistent to that port from the MLD Snooping forwarding table and applies the static or simulated joins again, until the number of IPv6 multicast groups joined by the port comes back within the configured threshold.*

Configuring IPv6 Multicast Group Replacement

For some special reasons, the number of IPv6 multicast groups passing through a switch or port may exceed the number configured for the switch or the port. In addition, in some specific applications, an IPv6 multicast group newly joined on the switch needs to replace an existing IPv6 multicast group automatically. A typical example is “channel switching”, namely, by joining the new multicast, a user automatically switches from the current IPv6 multicast group to the one.

To address this situation, you can enable the IPv6 multicast group replacement function on the switch or certain ports. When the number of IPv6 multicast groups a switch or a port has joined exceeds the limit.

- If the IPv6 multicast group replacement is enabled, the newly joined IPv6 multicast group automatically replaces an existing IPv6 multicast group with the lowest IPv6 address.

- If the IPv6 multicast group replacement is not enabled, new MLD reports will be automatically discarded.

Configuring IPv6 multicast group replacement globally

Follow these steps to configure IPv6 multicast group replacement globally:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter MLD Snooping view	mld-snooping	-
Configure IPv6 multicast group replacement	overflow-replace [vlan <i>vlan-list</i>]	Required Disabled by default

Configuring IPv6 multicast group replacement on a port or a group of ports

Follow these steps to configure IPv6 multicast group replacement on a port or a group of ports:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter the corresponding view	Enter Ethernet port view interface <i>interface-type interface-number</i>	Use either command
	Enter port group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	
Configure IPv6 multicast group replacement	mld-snooping overflow-replace [vlan <i>vlan-list</i>]	Required Disabled by default



CAUTION: Be sure to configure the maximum number of IPv6 multicast groups allowed on a port (refer to “Configuring Maximum Multicast Groups that that Can Be Joined on a Port” on page 594) before configuring IPv6 multicast group replacement. Otherwise, the IPv6 multicast group replacement functionality will not take effect.

Displaying and Maintaining MLD Snooping

To do...	Use the command...	Remarks
View the information about MLD Snooping multicast groups	display mld-snooping group [vlan <i>vlan-id</i>] [verbose]	Available in any view
View the statistics information of MLD messages learned by MLD Snooping	display mld-snooping statistics	Available in any view
Clear MLD Snooping multicast group information	reset mld-snooping group { <i>ipv6-group-address</i> all } [vlan <i>vlan-id</i>]	Available in user view
Clear the statistics information of all kinds of MLD messages learned by MLD Snooping	reset mld-snooping statistics	Available in user view



The **reset mld-snooping group** command cannot clear MLD Snooping multicast group information for static joins.

MLD Snooping Configuration Examples

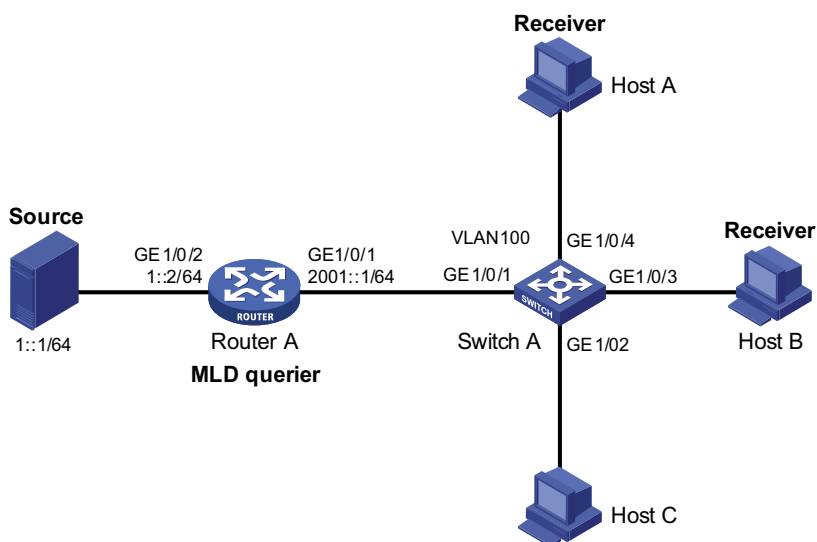
Simulated Joining Network requirements

As shown in Figure 177, Router A connects to the IPv6 multicast source through GigabitEthernet 1/0/2 and to Switch A through GigabitEthernet 1/0/1. Router A is the MLD querier on the subnet.

Perform the following configuration so that multicast data can be forwarded through GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 even if Host A and Host B temporarily stop receiving IPv6 multicast data for some unexpected reasons.

Network diagram

Figure 177 Network diagram for simulated joining configuration



Configuration procedure

- 1 Enable IPv6 forwarding and configure the IPv6 address of each interface
Enable IPv6 forwarding and configure an IPv6 address and prefix length for each interface as per Figure 177. The detailed configuration steps are omitted.
- 2 Configure Router A
Enable IPv6 multicast routing, enable IPv6 PIM-DM on each interface, and enable MLDv1 on GigabitEthernet 1/0/1.

```

<RouterA> system-view
[RouterA] multicast ipv6 routing-enable
[RouterA] interface GigabitEthernet 1/0/1
[RouterA-GigabitEthernet1/0/1] mld enable
[RouterA-GigabitEthernet1/0/1] pim ipv6 sm
[RouterA-GigabitEthernet1/0/1] quit
[RouterA] interface GigabitEthernet 1/0/2
[RouterA-GigabitEthernet1/0/2] pim ipv6 sm
[RouterA-GigabitEthernet1/0/2] quit
  
```

- 3 Configure Switch A

Enable MLD Snooping globally.

```
<SwitchA> system-view
[SwitchA] mld-snooping
[SwitchA-mld-snooping] quit
```

Create VLAN 100, assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/4 to this VLAN, and enable MLD Snooping in the VLAN.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port GigabitEthernet 1/0/1 to GigabitEthernet 1/0/4
[SwitchA-vlan100] mld-snooping enable
[SwitchA-vlan100] quit
```

Enable simulated host joining on GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4.

```
[SwitchA] interface GigabitEthernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] mld-snooping host-join ff1e::101 vlan 100
[SwitchA-GigabitEthernet1/0/3] quit
[SwitchA] interface GigabitEthernet 1/0/4
[SwitchA-GigabitEthernet1/0/4] mld-snooping host-join ff1e::101 vlan 100
[SwitchA-GigabitEthernet1/0/4] quit
```

4 Verify the configuration

View the detailed information about MLD Snooping multicast groups in VLAN 100 on Switch A.

```
[SwitchA] display mld-snooping group vlan 100 verbose
Total 1 IP Group(s).
Total 1 IP Source(s).
Total 1 MAC Group(s).

Port flags: D-Dynamic port, S-Static port, A-Aggregation port, C-Copy port
Subvlan flags: R-Real VLAN, C-Copy VLAN
Vlan(id):100.
Total 1 IP Group(s).
Total 1 IP Source(s).
Total 1 MAC Group(s).
Router port(s):total 1 port.
    GE1/0/1                (D) ( 00:01:30 )
IP group(s):the following ip group(s) match to one mac group.
IP group address:FF1E::101
( : , FF1E::101 ):
Attribute:      Host Port
Host port(s):total 2 port.
    GE1/0/3                (D) ( 00:03:23 )
    GE1/0/4                (D) ( 00:03:23 )
MAC group(s):
MAC group address:3333-0000-1001
Host port(s):total 2 port.
    GE1/0/3
    GE1/0/4
```

As shown above, GigabitEthernet 1/0/3 and GigabitEthernet 1/0/4 of Switch A have joined IPv6 multicast group FF1E::101.

Static Router Port Configuration

Network requirements

- As shown in Figure 178, Router A connects to an IPv6 multicast source (Source) through GigabitEthernet 1/0/2, and to Switch A through GigabitEthernet 1/0/1.
- MLD is to run on Router A, and MLD Snooping is to run on Switch A, Switch B and Switch C, with Router A acting as the MLD querier.

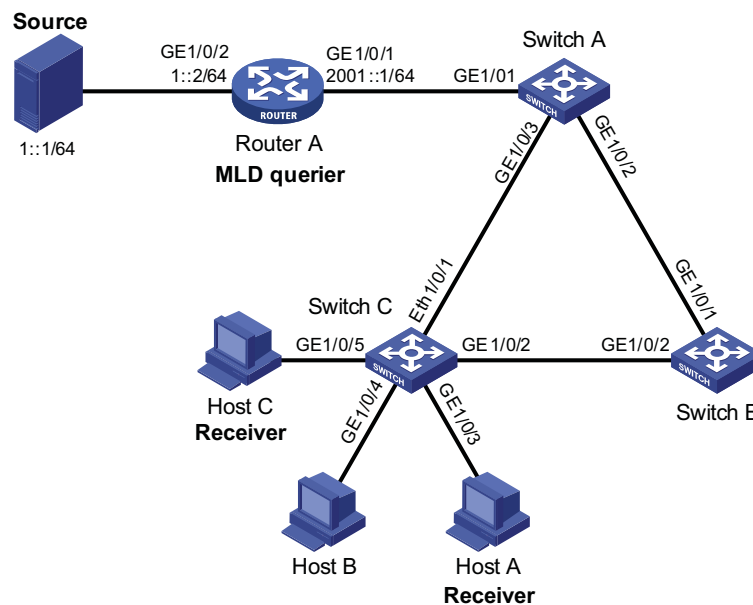
- Suppose STP runs on the network. To avoid data loops, the forwarding path from Switch A to Switch C is blocked under normal conditions, and IPv6 multicast traffic flows to the receivers, Host A and Host C, attached to Switch C, along the path of Switch A-Switch B-Switch C.
- Now it is required to configure GigabitEthernet 1/0/3 that connects Switch A to Switch C as a static router port, so that IPv6 multicast traffic can flow to the receivers nearly uninterruptedly along the path of Switch A-Switch C in the case that the path of Switch A-Switch B-Switch C gets blocked.



If no static router port is configured, when the path of Switch A-Switch B-Switch C gets blocked, at least one MLD query-response cycle must be completed before the IPv6 multicast data can flow to the receivers along the new path of Switch A-Switch C, namely IPv6 multicast delivery will be interrupted during this process.

Network diagram

Figure 178 Network diagram for static router port configuration



Configuration procedure

- 1 Enable IPv6 forwarding and configure the IPv6 address of each interface

Enable IPv6 forwarding and configure an IP address and prefix length for each interface as per Figure 178.

- 2 Configure Router A

Enable IPv6 multicast routing, enable IPv6 PIM-DM on each interface, and enable MLD on GigabitEthernet 1/0/1.

```
<RouterA> system-view
[RouterA] multicast ipv6 routing-enable
[RouterA] interface GigabitEthernet 1/0/1
[RouterA-GigabitEthernet 1/0/1] mld enable
[RouterA-GigabitEthernet 1/0/1] pim ipv6 dm
[RouterA-GigabitEthernet 1/0/1] quit
[RouterA] interface GigabitEthernet 1/0/2
```

```
[RouterA-GigabitEthernet 1/0/2] pim ipv6 dm
[RouterA-GigabitEthernet 1/0/2] quit
```

3 Configure Switch A

Enable MLD Snooping globally.

```
<SwitchA> system-view
[SwitchA] mld-snooping
[SwitchA-mld-snooping] quit
```

Create VLAN 100, assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to this VLAN, and enable MLD Snooping in the VLAN.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port GigabitEthernet 1/0/1 to GigabitEthernet 1/0/3
[SwitchA-vlan100] mld-snooping enable
[SwitchA-vlan100] quit
```

Configure GigabitEthernet 1/0/3 to be a static router port.

```
[SwitchA] interface GigabitEthernet 1/0/3
[SwitchA-GigabitEthernet 1/0/3] mld-snooping static-router-port vlan 100
[SwitchA-GigabitEthernet 1/0/3] quit
```

4 Configure Switch B

Enable MLD Snooping globally.

```
<SwitchB> system-view
[SwitchB] mld-snooping
[SwitchB-mld-snooping] quit
```

Create VLAN 100, assign GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to this VLAN, and enable MLD Snooping in the VLAN.

```
[SwitchB] vlan 100
[SwitchB-vlan100] port GigabitEthernet 1/0/1 GigabitEthernet 1/0/2
[SwitchB-vlan100] mld-snooping enable
[SwitchB-vlan100] quit
```

5 Configure Switch C

Enable MLD Snooping globally.

```
<SwitchC> system-view
[SwitchC] mld-snooping
[SwitchC-mld-snooping] quit
```

Create VLAN 100, assign GigabitEthernet 1/0/1 through GigabitEthernet 1/0/5 to this VLAN, and enable MLD Snooping in the VLAN.

```
[SwitchC] vlan 100
[SwitchC-vlan100] port GigabitEthernet 1/0/1 to GigabitEthernet 1/0/5
[SwitchC-vlan100] mld-snooping enable
[SwitchC-vlan100] quit
```

6 Verify the configuration

View the detailed information about MLD Snooping multicast groups in VLAN 100 on Switch A.

```
[SwitchA] display mld-snooping group vlan 100 verbose
    Total 1 IP Group(s).
    Total 1 IP Source(s).
    Total 1 MAC Group(s).
```

```
Port flags: D-Dynamic port, S-Static port, A-Aggregation port, C-Copy port
Subvlan flags: R-Real VLAN, C-Copy VLAN
Vlan(id):100.
```

```

Total 1 IP Group(s).
Total 1 IP Source(s).
Total 1 MAC Group(s).
Router port(s):total 2 port.
    GE1/0/1          (D) ( 00:01:30 )
    GE1/0/3          (S)
IP group(s):the following ip group(s) match to one mac group.
IP group address:FF1E::101
(::, FF1E::101):
Attribute:    Host Port
Host port(s):total 1 port.
    GE1/0/2          (D) ( 00:03:23 )
MAC group(s):
MAC group address:3333-0000-0101
Host port(s):total 1 port.
    GE1/0/2

```

As shown above, GigabitEthernet 1/0/3 of Switch A has become a static router port.

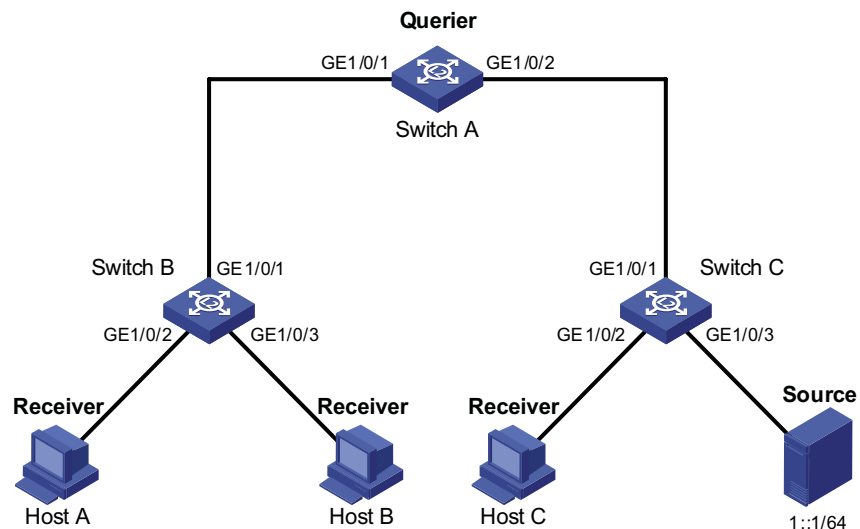
MLD Snooping Querier Configuration

Network requirements

- As shown in Figure 174, in a Layer-2-only network environment, Switch C is attached to the multicast source (Source) through GigabitEthernet 1/0/3. At least one receiver is connected to Switch B and Switch C respectively.
- MLDv1 is enabled on all the receivers. Switch A, Switch B, and Switch C run MLD Snooping. Switch A acts as the MLD Snooping querier.

Network diagram

Figure 179 Network diagram for MLD Snooping querier configuration



Configuration procedure

1 Configure switch A

Enable IPv6 forwarding and enable MLD Snooping globally.

```

<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] mld-snooping
[SwitchA-mld-snooping] quit

```

Create VLAN 100 and add GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to VLAN 100.

```
[SwitchA] vlan 100
[SwitchA-vlan100] port GigabitEthernet 1/0/1 GigabitEthernet 1/0/2
```

Enable MLD Snooping in VLAN 100 and configure the MLD-Snooping querier feature.

```
[SwitchA-vlan100] mld-snooping enable
[SwitchA-vlan100] mld-snooping querier
```

2 Configure Switch B

Enable IPv6 forwarding and enable MLD Snooping globally.

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] mld-snooping
[SwitchB-mld-snooping] quit
```

Create VLAN 100, add GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 into VLAN 100, and enable MLD Snooping in this VLAN.

```
[SwitchB] vlan 100
[SwitchB-vlan100] port GigabitEthernet 1/0/1 to GigabitEthernet 1/0/3
[SwitchB-vlan100] mld-snooping enable
```

3 Configuration on Switch C

Enable IPv6 forwarding and enable MLD Snooping globally.

```
<SwitchC> system-view
[SwitchC] ipv6
[SwitchC] mld-snooping
[SwitchC-mld-snooping] quit
```

Create VLAN 100, add GigabitEthernet 1/0/1 through GigabitEthernet 1/0/3 to VLAN 100, and enable MLD Snooping in this VLAN.

```
[SwitchC] vlan 100
[SwitchC-vlan100] port GigabitEthernet 1/0/1 to GigabitEthernet 1/0/3
[SwitchC-vlan100] mld-snooping enable
```

4 Verify the configuration

View the MLD message statistics on Switch C.

```
[SwitchC-vlan100] display mld-snooping statistics
Received MLD general queries:3.
Received MLDv1 specific queries:0.
Received MLDv1 reports:4.
Received MLD dones:0.
Sent      MLDv1 specific queries:0.
Received MLDv2 reports:0.
Received MLDv2 reports with right and wrong records:0.
Received MLDv2 specific queries:0.
Received MLDv2 specific sg queries:0.
Sent      MLDv2 specific queries:0.
Sent      MLDv2 specific sg queries:0.
Received error MLD messages:0.
```

Switch C received MLD general queries. This means that Switch A works as an MLD-Snooping querier.

Troubleshooting MLD Snooping

Switch Fails in Layer 2 Multicast Forwarding

Symptom

A switch fails to implement Layer 2 multicast forwarding.

Analysis

MLD Snooping is not enabled.

Solution

- 1 Enter the **display current-configuration** command to view the running status of MLD Snooping.
- 2 If MLD Snooping is not enabled, use the **mld-snooping** command to enable MLD Snooping globally, and then use **mld-snooping enable** command to enable MLD Snooping in VLAN view.
- 3 If MLD Snooping is disabled only for the corresponding VLAN, just use the **mld-snooping enable** command in VLAN view to enable MLD Snooping in the corresponding VLAN.

Configured IPv6 Multicast Group Policy Fails to Take Effect

Symptom

Although an IPv6 multicast group policy has been configured to allow hosts to join specific IPv6 multicast groups, the hosts can still receive IPv6 multicast data addressed to other groups.

Analysis

- The IPv6 ACL rule is incorrectly configured.
- The IPv6 multicast group policy is not correctly applied.
- The function of dropping unknown IPv6 multicast data is not enabled, so unknown IPv6 multicast data is flooded.
- Certain ports have been configured as static member ports of IPv6 multicasts groups, and this configuration conflicts with the configured IPv6 multicast group policy.

Solution

- 1 Use the **display acl ipv6** command to check the configured IPv6 ACL rule. Make sure that the IPv6 ACL rule conforms to the IPv6 multicast group policy to be implemented.
- 2 Use the **display this** command in MLD Snooping view or the corresponding interface view to check whether the correct IPv6 multicast group policy has been applied. If not, use the **group-policy** or **mld-snooping group-policy** command to apply the correct IPv6 multicast group policy.
- 3 Use the **display current-configuration** command to whether the function of dropping unknown IPv6 multicast data is enabled. If not, use the **mld-snooping drop-unknown** command to enable the function of dropping unknown IPv6 multicast data.
- 4 Use the **display mld-snooping group** command to check whether any port has been configured as a static member port of any IPv6 multicast group. If so, check

whether this configuration conflicts with the configured IPv6 multicast group policy. If any conflict exists, remove the port as a static member of the IPv6 multicast group.

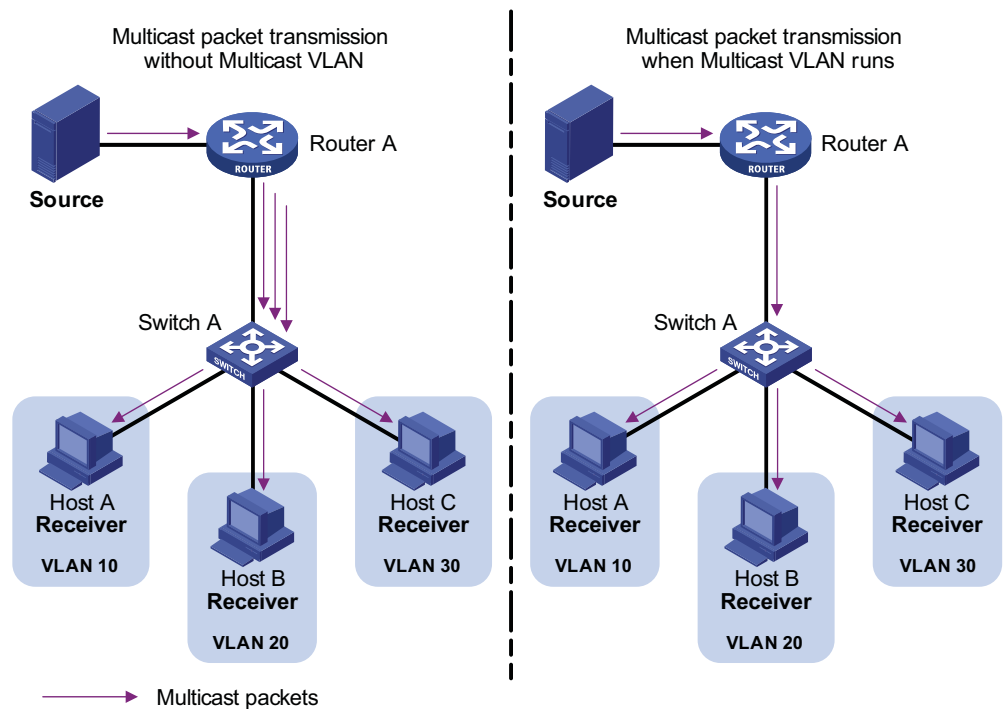
44

MULTICAST VLAN CONFIGURATION

Introduction to Multicast VLAN

As shown in Figure 180, in the traditional multicast programs-on-demand mode, when hosts that belong to different VLANs, Host A, Host B and Host C require multicast programs on demand service, Router A needs to forward a separate copy of the multicast data in each VLAN. This results in not only waste of network bandwidth but also extra burden on the Layer 3 device.

Figure 180 Before and after multicast VLAN is enabled on the Layer 2 device



To solve this problem, you can enable the multicast VLAN feature on Switch A, namely configure the VLANs to which these hosts belong as sub-VLANs of a multicast VLAN on the Layer 2 device and enable Layer 2 multicast in the multicast VLAN. After this configuration, Router A replicates the multicast data only within the multicast VLAN instead of forwarding a separate copy of the multicast data to each VLAN. This saves the network bandwidth and lessens the burden of the Layer 3 device.

Configuring Multicast VLAN

Follow these steps to configure a multicast VLAN:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Configure a specific VLAN as a multicast VLAN	multicast-vlan <i>vlan-id</i> enable	Required Disabled by default
Configure sub-VLANs for a specific multicast VLAN	multicast-vlan <i>vlan-id</i> subvlan <i>vlan-list</i>	Required No sub-VLAN by default.



- *The VLAN to be configured as the multicast VLAN and the VLANs to be configured as sub-VLANs of the multicast VLAN must exist.*
- *The number of sub-VLANs of the multicast VLAN must not exceed the system-defined limit (an Switch 4800G supports a maximum of one multicast VLAN and 127 sub-VLANs).*



CAUTION:

- *You cannot configure any multicast VLAN or a sub-VLAN of a multicast VLAN on a device with IP multicast routing or routing enabled.*
- *After a VLAN is configured into a multicast VLAN, IGMP Snooping must be enabled in the VLAN before the multicast VLAN feature can be implemented, while it is not necessary to enable IGMP Snooping in the sub-VLANs of the multicast VLAN.*

Displaying and Maintaining Multicast VLAN

To do...	Use the command...	Remarks
Display information about a multicast VLAN and its sub-VLANs	display multicast-vlan [<i>vlan-id</i>]	Available in any view

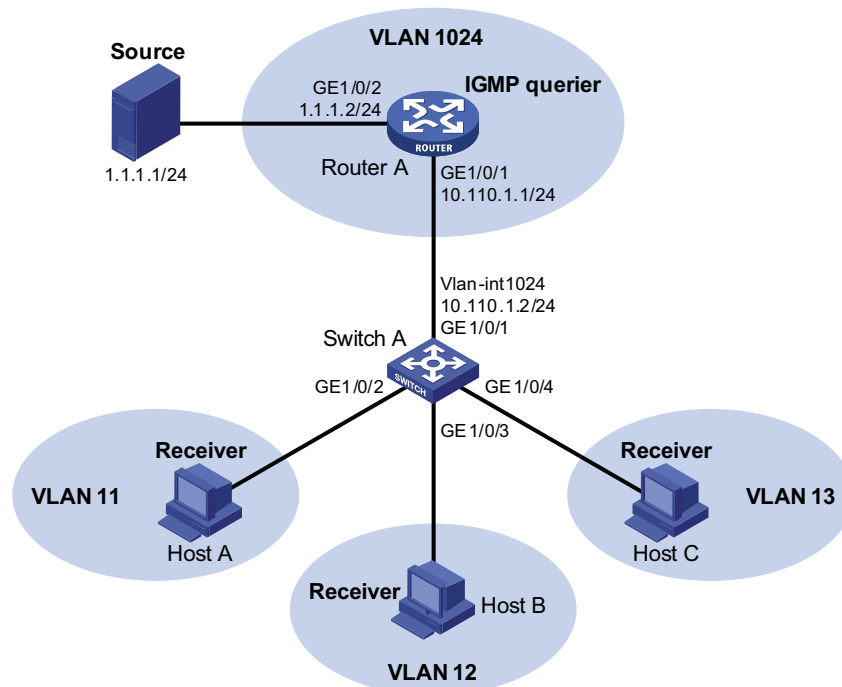
Multicast VLAN Configuration Example

Network requirements

- Router A connects to a multicast source through GigabitEthernet 1/0/2 and to Switch A, through GigabitEthernet 1/0/1.
- IGMP is required on Router A, and IGMP Snooping is required on Switch A. Router A is the IGMP querier.
- Switch A's GigabitEthernet 1/0/1 belongs to VLAN 1024, GigabitEthernet 1/0/2 through GigabitEthernet 1/0/4 belong to VLAN 11 through VLAN 13 respectively, and Host A through Host C are attached to GigabitEthernet 1/0/2 through GigabitEthernet 1/0/4 of Switch A.
- Configure the multicast VLAN feature so that Router A just sends multicast data to VLAN 1024 rather than to each VLAN when the three hosts attached to Switch A need the multicast data.

Network diagram

Figure 181 Network diagram for multicast VLAN configuration



Configuration procedure

1 Configure an IP address for each interconnecting interface

Configure an IP address and subnet mask for each interface as per Figure 181. The detailed configuration steps are omitted here.

2 Configure Router A

Enable IP multicast routing, enable PIM-DM on each interface and enable IGMP on GigabitEthernet 1/0/1.

```
<RouterA> system-view
[RouterA] multicast routing-enable
[RouterA] interface GigabitEthernet 1/0/1
[RouterA-GigabitEthernet 1/0/1] pim dm
[RouterA-GigabitEthernet 1/0/1] igmp enable
[RouterA-GigabitEthernet 1/0/1] quit
[RouterA] interface GigabitEthernet 1/0/2
[RouterA-GigabitEthernet 1/0/2] pim dm
[RouterA-GigabitEthernet 1/0/2] quit
```

3 Configure Switch A

Enable IGMP Snooping globally.

```
<SwitchA> system-view
[SwitchA] igmp-snooping
[SwitchA-igmp-snooping] quit
```

Create VLAN 11 and assign GigabitEthernet 1/0/2 to this VLAN.

```
[SwitchA] vlan 11
[SwitchA-vlan11] port GigabitEthernet 1/0/2
[SwitchA-vlan11] quit
```

The configuration for VLAN 12 and VLAN 13 is similar to the configuration for VLAN 11.

Create VLAN 1024, assign GigabitEthernet 1/0/1 to this VLAN and enable IGMP Snooping in the VLAN.

```
[SwitchA] vlan 1024
[SwitchA-vlan1024] port GigabitEthernet 1/0/1
[SwitchA-vlan1024] igmp-snooping enable
[SwitchA-vlan1024] quit
```

Configure VLAN 1024 as multicast VLAN and configure VLAN 11 through VLAN 13 as its sub-VLANs.

```
[SwitchA] multicast-vlan 1024 enable
[SwitchA] multicast-vlan 1024 subvlan 11 to 13
```

4 Verify the configuration

Display information about the multicast VLAN and its sub-VLANs.

```
[SwitchA] display multicast-vlan
multicast vlan 1024's subvlan list:
  Vlan 11-13
```

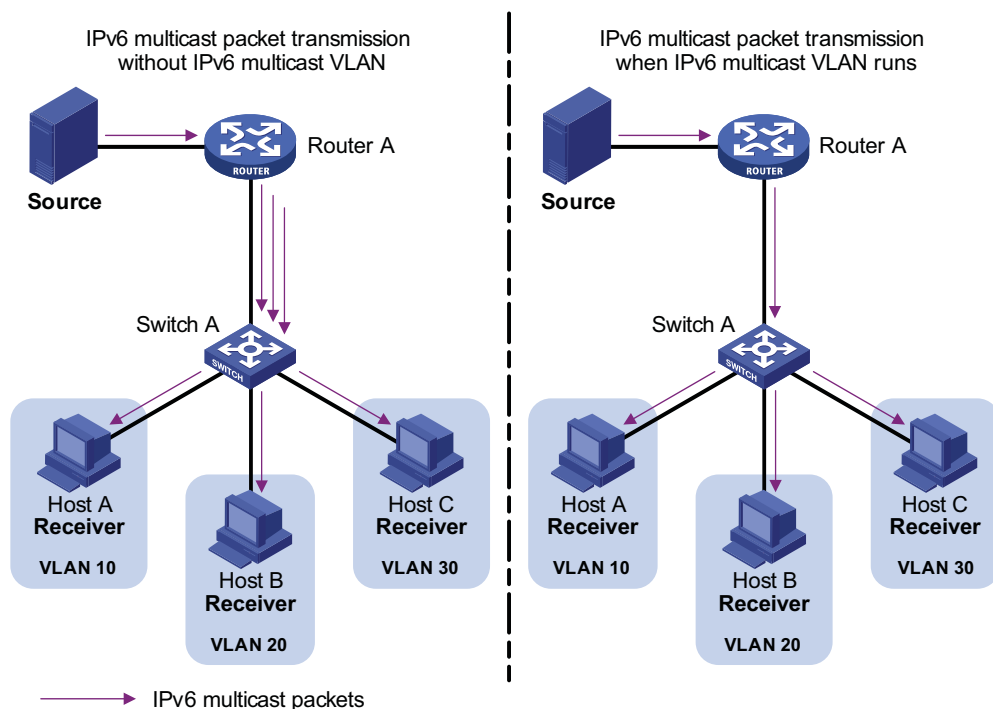
45

IPv6 MULTICAST VLAN CONFIGURATION

Introduction to IPv6 Multicast VLAN

As shown in Figure 182, in the traditional IPv6 multicast programs-on-demand mode, when hosts that belong to different VLANs, Host A, Host B and Host C require IPv6 multicast programs on demand service, Router A needs to forward a separate copy of the IPv6 multicast data in each VLAN. This results in not only waste of network bandwidth but also extra burden on the Layer 3 device.

Figure 182 Before and after IPv6 multicast VLAN is enabled on the Layer 2 device



To solve this problem, you can enable the IPv6 multicast VLAN feature on Switch A, namely configure the VLANs to which these hosts belong as sub-VLANs of an IPv6 multicast VLAN on the Layer 2 device and enable IPv6 Layer 2 multicast in the IPv6 multicast VLAN. After this configuration, Router A replicates the IPv6 multicast data only within the IPv6 multicast VLAN instead of forwarding a separate copy of the IPv6 multicast data to each VLAN. This saves the network bandwidth and lessens the burden of the Layer 3 device.

Configuring IPv6 Multicast VLAN

Follow these steps to configure IPv6 VLAN

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Configure a specific VLAN as an IPv6 multicast VLAN	multicast-vlan ipv6 <i>vlan-id</i> enable	Required By default, no VLAN is an IPv6 multicast VLAN.
Configure sub-VLANs for a multicast VLAN	multicast-vlan ipv6 <i>vlan-id</i> subvlan <i>vlan-list</i>	Required By default, no sub-VLANs exist.



- *The VLAN to be configured as an IPv6 multicast VLAN and the VLANs to be configured as sub-VLANs of the IPv6 multicast VLAN must exist.*
- *The total number of sub-VLANs of an IPv6 multicast VLAN must not exceed the system-defined limit (an Switch 4800G supports a maximum of one IPv6 multicast VLAN and 127 sub-VLANs).*



CAUTION:

- *You cannot enable IPv6 multicast VLAN on a device with IPv6 multicast routing enabled.*
- *After a VLAN is configured into an IPv6 multicast VLAN, MLD Snooping must be enabled in the VLAN before the IPv6 multicast VLAN feature can be implemented, while it is not necessary to enable MLD Snooping in the sub-VLANs of the IPv6 multicast VLAN.*

Displaying and Maintaining IPv6 Multicast VLAN

To do...	Use the command...	Remarks
Display information about an IPv6 multicast VLAN and its sub-VLANs	display multicast-vlan ipv6 [<i>vlan-id</i>]	Available in any view

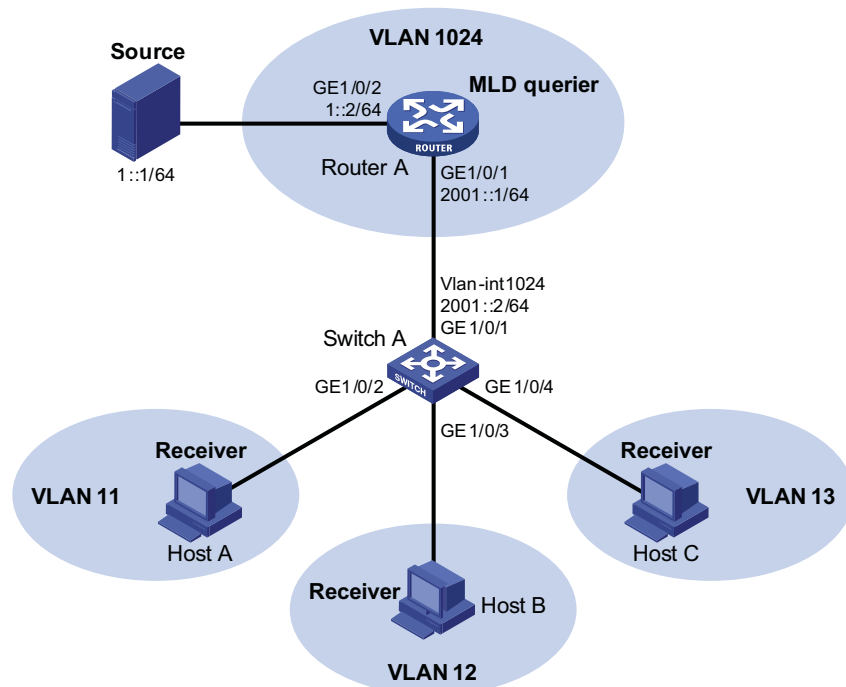
IPv6 Multicast VLAN Configuration Examples

Network requirements

- As shown in Figure 183, Router A connects to an IPv6 multicast source (Source) through GigabitEthernet 1/0/2, and to Switch A through GigabitEthernet 1/0/1.
- Router A is an IPv6 multicast router while Switch A is a Layer 2 switch. Router A acts as the MLD querier on the subnet.
- Switch A's GigabitEthernet 1/0/1 belongs to VLAN 1024, GigabitEthernet 1/0/2 through GigabitEthernet 1/0/4 belong to VLAN 11 through VLAN 13 respectively, and Host A through Host C are attached to GigabitEthernet 1/0/2 through GigabitEthernet 1/0/4 of Switch A.
- Configure the IPv6 multicast VLAN feature so that Router A just sends IPv6 multicast data to VLAN 1024 rather than to each VLAN when the three hosts attached to Switch A need the IPv6 multicast data.

Network diagram

Figure 183 Network diagram for IPv6 multicast VLAN configuration



Configuration procedure

- 1 Enable IPv6 forwarding and configure IPv6 addresses of the interfaces of each device.

Enable IPv6 forwarding and configure the IPv6 address and address prefix for each interface as per Figure 183. The detailed configuration steps are omitted here.

- 2 Configure Router A

Enable IPv6 multicast routing, enable IPv6 PIM-DM on each interface, and enable MLD on GigabitEthernet 1/0/1.

```
<RouterA> system-view
[RouterA] multicast ipv6 routing-enable
[RouterA] interface GigabitEthernet 1/0/1
[RouterA-GigabitEthernet1/0/1] pim ipv6 dm
[RouterA-GigabitEthernet1/0/1] mld enable
[RouterA-GigabitEthernet1/0/1] quit
[RouterA] interface GigabitEthernet1/0/2
[RouterA-GigabitEthernet1/0/2] pim ipv6 dm
[RouterA-GigabitEthernet1/0/2] quit
```

- 3 Configure Switch A

Enable MLD Snooping globally.

```
<SwitchA> system-view
[SwitchA] mld-snooping
[SwitchA-mld-snooping] quit
```

Create VLAN 11 and add GigabitEthernet 1/0/2 into VLAN 11.

```
[SwitchA] vlan 11
[SwitchA-vlan11] port GigabitEthernet 1/0/2
[SwitchA-vlan11] quit
```

The configuration for VLAN 12 and VLAN 13 is similar. The detailed configuration steps are omitted.

Create VLAN 1024, add GigabitEthernet 1/0/1 to VLAN 1024, and enable MLD Snooping in this VLAN.

```
[SwitchA] vlan 1024
[SwitchA-vlan1024] port GigabitEthernet 1/0/1
[SwitchA-vlan1024] mld-snooping enable
[SwitchA-vlan1024] quit
```

Configure VLAN 1024 as an IPv6 multicast VLAN, and configure VLAN 11 through VLAN 13 as its sub-VLANs.

```
[SwitchA] multicast-vlan ipv6 1024 enable
[SwitchA] multicast-vlan ipv6 1024 subvlan 11 to 13
```

4 Verify the configuration

Display IPv6 multicast VLAN and sub-VLAN information on Switch A.

```
[SwitchA] display multicast-vlan ipv6
IPv6 multicast vlan 1024's subvlan list:
  vlan 11-13
```


46

IGMP CONFIGURATION

When configuring IGMP, go to the following sections for the information you are interested in:

- "IGMP Overview" on page 613
- "IGMP Configuration Task List" on page 617
- "IGMP Configuration Example" on page 624
- "Troubleshooting IGMP" on page 626



The term "router" in this document refers to a router in a generic sense or a Layer 3 switch running IGMP.

IGMP Overview

As a TCP/IP protocol responsible for IP multicast group member management, the Internet Group Management Protocol (IGMP) is used by IP hosts to establish and maintain their multicast group memberships to immediately neighboring multicast routers.

IGMP Versions

So far, there are three IGMP versions:

- IGMPv1 (documented in RFC 1112)
- IGMPv2 (documented in RFC 2236)
- IGMPv3 (documented in RFC 3376)

All IGMP versions support the Any-Source Multicast (ASM) model. In addition, IGMPv3 can be directly used to implement the Source-Specific Multicast (SSM) model.

Work Mechanism of IGMPv1

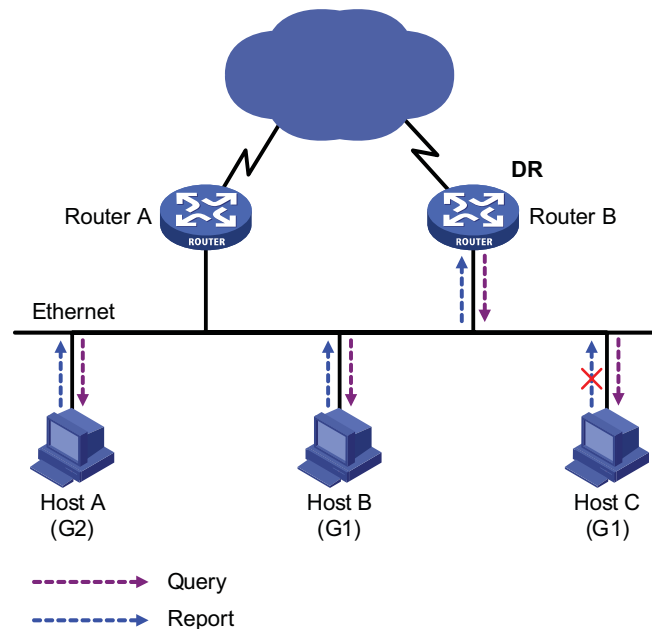
IGMPv1 manages multicast group memberships mainly based on the query and response mechanism.

Of multiple multicast routers on the same subnet, all the routers can hear IGMP membership report messages (often referred to as reports) from hosts, but only one router is needed for sending IGMP query messages (often referred to as queries). So, a querier election mechanism is required to determine which router will act as the IGMP querier on the subnet.

In IGMPv1, the designated router (DR) elected by a multicast routing protocol (such as PIM) serves as the IGMP querier.



For more information about DR, refer to "DR election" on page 633.

Figure 184 Joining multicast groups

Assume that Host B and Host C are expected to receive multicast data addressed to multicast group G1, while Host A is expected to receive multicast data addressed to G2, as shown in Figure 184. The basic process that the hosts join the multicast groups is as follows:

- 1 The IGMP querier (Router B in the figure) periodically multicasts IGMP queries (with the destination address of 224.0.0.1) to all hosts and routers on the local subnet.
- 2 Upon receiving a query message, Host B or Host C (the delay timer of whichever expires first) sends an IGMP report to the multicast group address of G1, to announce its interest in G1. Assume it is Host B that sends the report message.
- 3 Host C, which is on the same subnet, hears the report from Host B for joining G1. Upon hearing the report, Host C will suppress itself from sending a report message for the same multicast group, because the IGMP routers (Router A and Router B) already know that at least one host on the local subnet is interested in G1. This mechanism, known as IGMP report suppression, helps reduce traffic over the local subnet.
- 4 At the same time, because Host A is interested in G2, it sends a report to the multicast group address of G2.
- 5 Through the above-mentioned query/report process, the IGMP routers learn that members of G1 and G2 are attached to the local subnet, and generate (*, G1) and (*, G2) multicast forwarding entries, which will be the basis for subsequent multicast forwarding, where * represents any multicast source.
- 6 When the multicast data addressed to G1 or G2 reaches an IGMP router, because the (*, G1) and (*, G2) multicast forwarding entries exist on the IGMP router, the router forwards the multicast data to the local subnet, and then the receivers on the subnet receive the data.

As IGMPv1 does not specifically define a Leave Group message, upon leaving a multicast group, an IGMPv1 host stops sending reports with the destination

address being the address of that multicast group. If no member of a multicast group exists on the subnet, the IGMP routers will not receive any report addressed to that multicast group, so the routers will delete the multicast forwarding entries corresponding to that multicast group after a period of time.

Enhancements Provided by IGMPv2

Compared with IGMPv1, IGMPv2 provides the querier election mechanism and Leave Group mechanism.

Querier election mechanism

In IGMPv1, the DR elected by the Layer 3 multicast routing protocol (such as PIM) serves as the querier among multiple routers on the same subnet.

In IGMPv2, an independent querier election mechanism is introduced. The querier election process is as follows:

- 1 Initially, every IGMPv2 router assumes itself as the querier and sends IGMP general query messages (often referred to as general queries) to all hosts and routers on the local subnet (the destination address is 224.0.0.1).
- 2 Upon hearing a general query, every IGMPv2 router compares the source IP address of the query message with its own interface address. After comparison, the router with the lowest IP address wins the querier election and all other IGMPv2 routers become non-queriers.
- 3 All the non-queriers start a timer, known as "other querier present timer". If a router receives an IGMP query from the querier before the timer expires, it resets this timer; otherwise, it assumes the querier to have timed out and initiates a new querier election process.

Leave group" mechanism

In IGMPv1, when a host leaves a multicast group, it does not send any notification to the multicast router. The multicast router relies on host response timeout to know whether a group no longer has members. This adds to the leave latency.

In IGMPv2, on the other hand, when a host leaves a multicast group:

- 1 This host sends a Leave Group message (often referred to as leave message) to all routers (the destination address is 224.0.0.2) on the local subnet.
- 2 Upon receiving the leave message, the querier sends a configurable number of group-specific queries to the group being left. The destination address field and group address field of the message are both filled with the address of the multicast group being queried.
- 3 One of the remaining members, if any on the subnet, of the group being queried should send a membership report within the maximum response time set in the query messages.
- 4 If the querier receives a membership report for the group within the maximum response time, it will maintain the memberships of the group; otherwise, the querier will assume that no hosts on the subnet are still interested in multicast traffic to that group and will stop maintaining the memberships of the group.

Enhancements in IGMPv3



The support for the Exclude mode varies with device models.

Built upon and being compatible with IGMPv1 and IGMPv2, IGMPv3 provides hosts with enhanced control capabilities and provides enhancements of query and report messages.

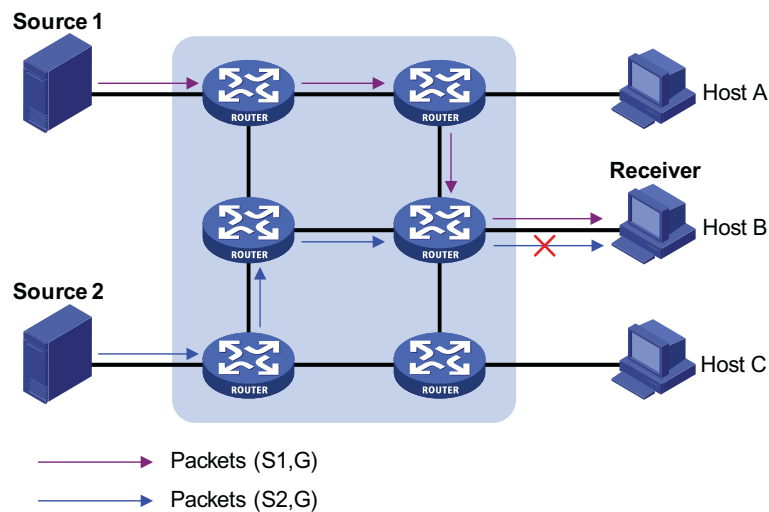
Enhancements in control capability of hosts

IGMPv3 has introduced source filtering modes (Include and Exclude), so that a host not only can join a designated multicast group but also can specify to receive or reject multicast data from a designated multicast source. When a host joins a multicast group:

- If it needs to receive multicast data from specific sources like S1, S2, ..., it sends a report with the Filter-Mode denoted as "Include Sources (S1, S2,).
- If it needs to reject multicast data from specific sources like S1, S2, ..., it sends a report with the Filter-Mode denoted as "Exclude Sources (S1, S2,).

As shown in Figure 185, the network comprises two multicast sources, Source 1 (S1) and Source 2 (S2), both of which can send multicast data to multicast group G. Host B is interested only in the multicast data that Source 1 sends to G but not in the data from Source 2.

Figure 185 Flow paths of source-and-group-specific multicast traffic



In the case of IGMPv1 or IGMPv2, Host B cannot select multicast sources when it joins multicast group G. Therefore, multicast streams from both Source 1 and Source 2 will flow to Host B whether it needs them or not.

When IGMPv3 is running between the hosts and routers, Host B can explicitly express its interest in the multicast data Source 1 sends to multicast group G (denoted as (S1, G)), rather than the multicast data Source 2 sends to multicast group G (denoted as (S2, G)). Thus, only multicast data from Source 1 will be delivered to Host B.

Enhancements in query and report capabilities

1 Query message carrying the source addresses

IGMPv3 supports not only general queries (feature of IGMPv1) and group-specific queries (feature of IGMPv2), but also group-and-source-specific queries.

- A general query does not carry a group address, nor a source address;
- A group-specific query carries a group address, but no source address;
- A group-and-source-specific query carries a group address and one or more source addresses.

2 Reports containing multiple group records

Unlike an IGMPv1 or IGMPv2 report message, an IGMPv3 report message is destined to 224.0.0.22 and contains one or more group records. Each group record contains a multicast group address and a multicast source address list.

Group record types include:

- IS_IN: The source filtering mode is Include, namely, the report sender requests the multicast data from only the sources defined in the specified multicast source list. If the specified multicast source list is empty, this means that the report sender has left the reported multicast group.
- IS_EX: The source filtering mode is Exclude, namely, the report sender requests the multicast data from any sources but those defined in the specified multicast source list.
- TO_IN: The filter mode has changed from Exclude to Include.
- TO_EX: The filter mode has changed from Include to Exclude.
- ALLOW: The Source Address fields in this Group Record contain a list of the additional sources that the system wishes to hear from, for packets sent to the specified multicast address. If the change was to an Include source list, these are the addresses that were added to the list; if the change was to an Exclude source list, these are the addresses that were deleted from the list.
- BLOCK: indicates that the Source Address fields in this Group Record contain a list of the sources that the system no longer wishes to hear from, for packets sent to the specified multicast address. If the change was to an Include source list, these are the addresses that were deleted from the list; if the change was to an Exclude source list, these are the addresses that were added to the list.

Protocols and Standards The following documents describe different IGMP versions:

- RFC 1112: Host Extensions for IP Multicasting
- RFC 2236: Internet Group Management Protocol, Version 2
- RFC 3376: Internet Group Management Protocol, Version 3

IGMP Configuration Task List

Complete these tasks to configure IGMP:

Task	Description	
"Configuring Basic Functions of IGMP" on page 618	"Enabling IGMP" on page 618	Required
	"Configuring IGMP Versions" on page 619	Optional
	"Configuring a Static Member of a Multicast Group" on page 619	Optional
	"Configuring a Multicast Group Filter" on page 620	Optional
"Adjusting IGMP Performance" on page 620	"Configuring IGMP Message Options" on page 620	Optional
	"Configuring IGMP Query and Response Parameters" on page 621	Optional
	"Configuring IGMP Fast Leave Processing" on page 623	Optional



- *Configurations performed in IGMP view are effective on all interfaces, while configurations performed in interface view are effective on the current interface only.*
- *If a feature is not configured for an interface in interface view, the global configuration performed in IGMP view will apply to that interface. If a feature is configured in both IGMP view and interface view, the configuration performed in interface view will be given priority.*

Configuring Basic Functions of IGMP

Configuration Prerequisites

Before configuring the basic functions of IGMP, complete the following tasks:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Configure PIM-DM or PIM-SM

Before configuring the basic functions of IGMP, prepare the following data:

- IGMP version
- Multicast group and multicast source addresses for static group member configuration
- ACL rule for multicast group filtering

Enabling IGMP

First, IGMP must be enabled on the interface on which the multicast group memberships are to be established and maintained.

Follow these steps to enable IGMP:

To do...	Use the command...	Description
Enter system view	system-view	-

To do...	Use the command...	Description
Enable IP multicast routing	multicast routing-enable	Required Disabled by default
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable IGMP	igmp enable	Required Disabled by default

Configuring IGMP Versions

Because messages vary with different IGMP versions, the same IGMP version should be configured for all routers on the same subnet before IGMP can work properly.

Configuring an IGMP version globally

Follow these steps to configure an IGMP version globally:

To do...	Use the command...	Description
Enter system view	system-view	-
Enter IGMP view	igmp	-
Configure an IGMP version globally	version <i>version-number</i>	Optional IGMPv2 by default

Configuring an IGMP version on an interface

Follow these steps to configure an IGMP version on an interface:

To do...	Use the command...	Description
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure an IGMP version on the interface	igmp version <i>version-number</i>	Optional IGMPv2 by default

Configuring a Static Member of a Multicast Group

After an interface is configured as a static member of a multicast group, it will act as a virtual member of the multicast group to receive multicast data addressed to that multicast group for the purpose of testing multicast data forwarding.

Follow these steps to configure an interface as a statically connected member of a multicast group:

To do...	Use the command...	Description
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure the interface as a static member of a multicast group	igmp static-group <i>group-address</i> [source <i>source-address</i>]	Required An interface is not a static member of any multicast group by default.



- Before you can configure an interface of a PIM-SM device as a static member of a multicast group, if the interface is PIM-SM enabled, it must be a PIM-SM DR; if this interface is IGMP enabled but not PIM-SM enabled, it must be an IGMP querier.
- As a static member of a multicast group, an interface does not respond to the queries from the IGMP querier, nor does it send an unsolicited IGMP membership report or an IGMP leave group message when it joins or leaves a multicast group. In other words, the interface will not become a real member of the multicast group.

Configuring a Multicast Group Filter

You can configure a multicast group filter in IGMP Snooping. For details, see “Configuring a Multicast Group Filter” on page 566.

Adjusting IGMP Performance



For the configuration tasks described in this section

- Configurations performed in IGMP view are effective on all interfaces, while configurations performed in interface view are effective on the current interface only.
- If the same feature is configured in both IGMP view and interface view, the configuration performed in interface view is given priority, regardless of the configuration sequence.

Configuration Prerequisites

Before adjusting IGMP performance, complete the following tasks:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Configure basic functions of IGMP

Before adjusting IGMP performance, prepare the following data:

- IGMP general query interval
- IGMP querier’s robustness variable
- Maximum response time for IGMP general queries
- IGMP last-member query interval
- Other querier present interval

Configuring IGMP Message Options

As IGMPv2 and IGMPv3 involve group-specific and group-and-source-specific queries, and multicast groups change dynamically, a device cannot join all multicast groups. Therefore, when receiving a multicast packet but unable to locate the outgoing interface for the destination multicast group, an IGMP router needs to leverage the Router-Alert option to pass the multicast packet to the upper-layer protocol for processing. For details about the Router-Alert option, refer to RFC 2113.

An IGMP message is processed differently depending whether it carries the Router-Alert option in the IP header:

- By default, for the consideration of compatibility, the device does not check the Router-Alert option, namely it processes all the IGMP messages it received. In this case, IGMP messages are directly passed to the upper layer protocol, no matter whether the IGMP messages carry the Router-Alert option or not.
- To enhance the device performance and avoid unnecessary costs, and also for the consideration of protocol security, you can configure the device to discard IGMP messages that do not carry the Router-Alert option.

Configuring IGMP packet options globally

Follow these steps to configure IGMP packet options globally:

To do...	Use the command...	Description
Enter system view	system-view	-
Enter IGMP view	igmp	-
Configure the router to discard any IGMP message that does not carry the Router-Alert option	require-router-alert	Optional By default, the device does not check the Router-Alert option.
Enable the insertion of the Router-Alert option into IGMP messages	send-router-alert	Optional By default, IGMP messages carry the Router-Alert option.

Configuring IGMP packet options on an interface

Follow these steps to configure IGMP packet options on an interface:

To do...	Use the command...	Description
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure the interface to discard any IGMP message that does not carry the Router-Alert option	igmp require-router-alert	Optional By default, the device does not check the Router-Alert option.
Enable the insertion of the Router-Alert option into IGMP messages	igmp send-router-alert	Optional By default, IGMP messages carry the Router-Alert option.

Configuring IGMP Query and Response Parameters

The IGMP querier periodically sends IGMP general queries at the "IGMP query interval" to determine whether any multicast group member exists on the network. You can tune the IGMP general query interval based on actual condition of the network.

On startup, the IGMP querier sends "startup query count" IGMP general queries at the "startup query interval", which is 1/4 of the "IGMP query interval". Upon receiving an IGMP leave message, the IGMP querier sends "last member query count" IGMP group-specific queries at the "IGMP last member query interval". Both startup query count and last member query count are set to the IGMP querier robustness variable.

IGMP is robust to “robustness variable minus 1” packet losses on a network. Therefore, a greater value of the robustness variable makes the IGMP querier “more robust”, but results in a longer multicast group timeout time.

Upon receiving an IGMP query (general query or group-specific query), a host starts a delay timer for each multicast group it has joined. This timer is initialized to a random value in the range of 0 to the maximum response time, which is derived from the Max Response Time field in the IGMP query. When the timer value comes down to 0, the host sends an IGMP report to the corresponding multicast group.

An appropriate setting of the maximum response time for IGMP queries allows hosts to respond to queries quickly and avoids bursts of IGMP traffic on the network caused by reports simultaneously sent by a large number of hosts when the corresponding timers expires simultaneously.

- For IGMP general queries, you can configure the maximum response time to fill their Max Response time field.
- For IGMP group-specific queries, you can configure the IGMP last member query interval to fill their Max Response time field. Namely, for IGMP group-specific queries, the maximum response time equals the IGMP last member query interval.

When multiple multicast routers exist on the same subnet, the IGMP querier is responsible for sending IGMP queries. If a non-querier router receives no IGMP query from the querier within the “other querier present interval”, it will assume the querier to have expired and a new querier election process is launched; otherwise, the non-querier router will reset its “other querier present timer”.

Configuring IGMP query and response parameters globally

Follow these steps to configure IGMP query and response parameters globally:

To do...	Use the command...	Description
Enter system view	system-view	-
Enter IGMP view	igmp	-
Configure the IGMP query interval	timer query <i>interval</i>	Optional 60 seconds by default
Configure the IGMP querier robustness variable	robust-count <i>robust-value</i>	Optional 2 by default
Configure the maximum response time for IGMP general queries	max-response-time <i>interval</i>	Optional 10 seconds by default
Configure the IGMP last member query interval	Last-member-query-interval <i>interval</i>	Optional 1 second by default
Configure the other querier present interval	timer other-querier-present <i>interval</i>	Optional For the system default, see “Note” below.

Configuring IGMP query and response parameters on an interface

Follow these steps to configure IGMP query and response parameters on an interface:

To do...	Use the command...	Description
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure IGMP query interval	igmp timer query <i>interval</i>	Optional 60 seconds by default
Configure the IGMP querier robustness variable	igmp robust-count <i>robust-value</i>	Optional 2 by default
Configure the maximum response time for IGMP general queries	igmp max-response-time <i>interval</i>	Optional 10 seconds by default
Configure the IGMP last member query interval	igmp last-member-query-interval <i>interval</i>	Optional 1 second by default
Configure the other querier present interval	igmp timer other-querier-present <i>interval</i>	Optional For the system default, see "Note" below.



- *If not statically configured, the other querier present interval is [IGMP query interval] times [IGMP robustness variable] plus [maximum response time for IGMP general queries] divided by two. By default, the values of these three parameters are 60 (seconds), 2 and 10 (seconds) respectively, so the default value of the other querier present interval = $60 \times 2 + 10 / 2 = 125$ (seconds).*
- *If statically configured, the other querier present interval takes the configured value.*



CAUTION:

- *Make sure that the other querier present interval is greater than the IGMP query interval; otherwise the IGMP querier may change frequently on the network.*
- *Make sure that the IGMP query interval is greater than the maximum response time for IGMP general queries; otherwise, multicast group members may be wrongly removed.*
- *The configurations of the maximum response time for IGMP general queries, the IGMP last member query interval and the IGMP other querier present interval are effective only for IGMPv2 or IGMPv3.*

Configuring IGMP Fast Leave Processing

Fast leave processing is implemented by IGMP Snooping. For details, see "Configuring Fast Leave Processing" on page 562.

Displaying and Maintaining IGMP

To do...	Use the command...	Description
View IGMP multicast group information	display igmp group [<i>group-address</i> interface <i>interface-type</i> <i>interface-number</i>] [static verbose]	Available in any view
View IGMP layer 2 port information	display igmp group port-info [<i>vlan</i> <i>vlan-id</i>] [verbose]	Available in any view

To do...	Use the command...	Description
View IGMP configuration and running information	display igmp interface [<i>interface-type interface-number</i>] [verbose]	Available in any view
View routing information in the IGMP routing table	display igmp routing-table [<i>source-address</i> [mask { <i>mask</i> <i>mask-length</i> }]] <i>group-address</i> [mask { <i>mask</i> <i>mask-length</i> }]] *	Available in any view
Clear IGMP forwarding entries	reset igmp group { all interface <i>interface-type interface-number</i> } { all <i>group-address</i> [mask { <i>mask</i> <i>mask-length</i> }] [<i>source-address</i> [mask { <i>mask</i> <i>mask-length</i> }]] }	Available in user view
Clear Layer 2 port information about IGMP multicast groups	reset igmp group port-info { all <i>group-address</i> } [vlan <i>vlan-id</i>]	Available in user view



- The **reset igmp group** command cannot clear the IGMP forwarding entries of static joins.
- The **reset igmp group port-info** command cannot clear Layer 2 port information about IGMP multicast groups of static joins.



CAUTION: The **reset igmp group** command may cause an interruption of receivers' reception of multicast data.

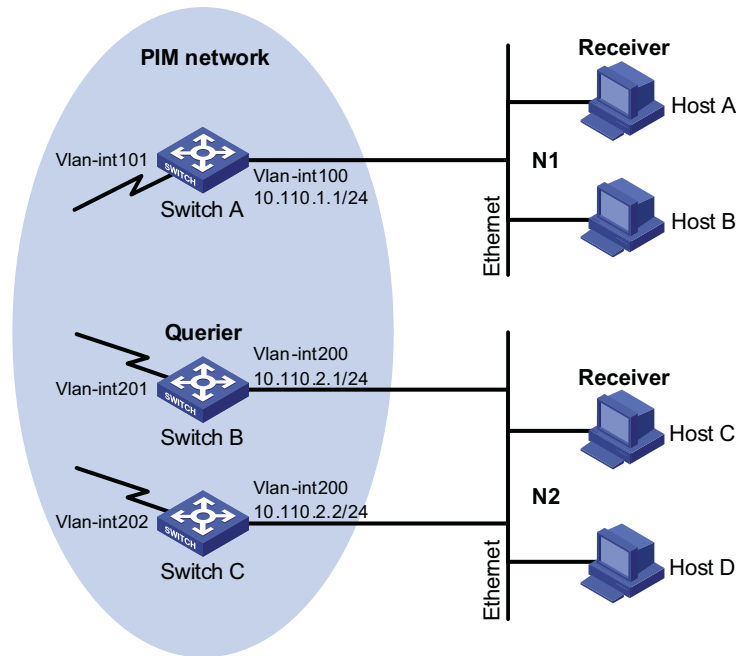
IGMP Configuration Example

Network requirements

- Receivers receive VOD information through the multicast mode. Receivers of different organizations form stub networks N1 and N2, and Host A and Host C are receivers in N1 and N2 respectively.
- Switch A in the PIM network connects to N1, and both Switch B and Switch C connect to N2.
- Switch A connects to N1 through VLAN-interface 100, and to other devices in the PIM network through VLAN-interface 101.
- Switch B and Switch C connect to N2 through their respective VLAN-interface 200, and to other devices in the PIM network through VLAN-interface 201 and VLAN-interface 202 respectively.
- IGMPv3 is required between Switch A and N1. IGMPv2 is required between the other two switches and N2, with Switch B as the IGMP querier.

Network diagram

Figure 186 Network diagram for IGMP configuration



Configuration procedure

- 1 Configure the IP addresses of the switch interfaces and configure a unicast routing protocol

Configure the IP address and subnet mask of each interface as per Figure 186. The detailed configuration steps are omitted here.

Configure the OSPF protocol for interoperability among the switches. Ensure the network-layer interoperability among Switch A, Switch B and Switch C on the PIM network and dynamic update of routing information among the switches through a unicast routing protocol. The detailed configuration steps are omitted here.

- 2 Enable IP multicast routing, and enable IGMP on the host-side interfaces

Enable IP multicast routing on Switch A, and enable IGMP (version 3) on VLAN-interface 100.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] igmp enable
[SwitchA-Vlan-interface100] igmp version 3
[SwitchA-Vlan-interface100] quit
```

Enable IP multicast routing on Switch B, and enable IGMP (version 2) on VLAN-interface 200.

```
<SwitchB> system-view
[SwitchB] multicast routing-enable
[SwitchB] interface vlan-interface 200
[SwitchB-Vlan-interface200] igmp enable
[SwitchB-Vlan-interface200] igmp version 2
[SwitchB-Vlan-interface200] quit
```

Enable IP multicast routing on Switch C, and enable IGMP (version 2) on VLAN-interface 200.

```
<SwitchC> system-view
[SwitchC] multicast routing-enable
[SwitchC] interface vlan-interface 200
[SwitchC-Vlan-interface200] igmp enable
[SwitchC-Vlan-interface200] igmp version 2
[SwitchC-Vlan-interface200] quit
```

3 Verify the configuration

Carry out the **display igmp interface** command to view the IGMP configuration and running status on each switch interface. For example:

View IGMP information on VLAN-interface 200 of Switch B.

```
[SwitchB] display igmp interface vlan-interface 200
Vlan-interface200(10.110.2.1):
  IGMP is enabled
  Current IGMP version is 2
  Value of query interval for IGMP(in seconds): 60
  Value of other querier timeout for IGMP(in seconds): 125
  Value of maximum query response time for IGMP(in seconds): 10
  Querier for IGMP: 10.110.2.1 (this router)
  Total 1 IGMP Group reported
```

Troubleshooting IGMP

No Member Information on the Receiver-Side Router

Symptom

When a host sends a report for joining multicast group G, there is no member information of the multicast group G on the router closest to that host.

Analysis

- The correctness of networking and interface connections directly affects the generation of group member information.
- Multicast routing must be enabled on the router.
- If the **igmp group-policy** command has been configured on the interface, the interface cannot receive report messages that fail to pass filtering.

Solution

- 1 Check that the networking is correct and interface connections are correct.
- 2 Check that the interfaces and the host are on the same subnet. Use the **display current-configuration interface** command to view the IP address of the interface.
- 3 Check that multicast routing is enabled. Carry out the **display current-configuration** command to check whether the **multicast routing-enable** command has been executed. If not, carry out the **multicast routing-enable** command in system view to enable IP multicast routing. In addition, check that IGMP is enabled on the corresponding interfaces.
- 4 Check that the interface is in normal state and the correct IP address has been configured. Carry out the **display igmp interface** command to view the interface information. If no interface information is output, this means the interface is

abnormal. Typically this is because the **shutdown** command has been executed on the interface, or the interface connection is incorrect, or no correct IP address has been configured on the interface.

- 5 Check that no ACL rule has been configured to restrict the host from joining the multicast group G. Carry out the **display current-configuration interface** command to check whether the **igmp group-policy** command has been executed. If the host is restricted from joining the multicast group G, the ACL rule must be modified to allow receiving the reports for the multicast group G.

Inconsistent Memberships on Routers on the Same Subnet

Symptom

Different memberships are maintained on different IGMP routers on the same subnet.

Analysis

- A router running IGMP maintains multiple parameters for each interface, and these parameters influence one another, forming very complicated relationships. Inconsistent IGMP interface parameter configurations for routers on the same subnet will surely result in inconsistency of memberships.
- In addition, although IGMP routers are compatible with hosts, all routers on the same subnet must run the same version of IGMP. Inconsistent IGMP versions running on routers on the same subnet will also lead to inconsistency of IGMP memberships.

Solution

- 1 Check the IGMP configuration. Carry out the **display current-configuration** command to view the IGMP configuration information on the interfaces.
- 2 Carry out the **display igmp interface** command on all routers on the same subnet to check the IGMP-related timer settings. Make sure that the settings are consistent on all the routers.
- 3 Use the **display igmp interface** command to check whether the routers are running the same version of IGMP.

When configuring PIM, go to these sections for information you are interested in:

- “PIM Overview” on page 629
- “Configuring PIM-DM” on page 641
- “Configuring PIM-SM” on page 643
- “Configuring PIM-SSM” on page 652
- “Configuring PIM Common Information” on page 653
- “Displaying and Maintaining PIM” on page 658
- “PIM Configuration Examples” on page 659
- “Troubleshooting PIM Configuration” on page 669



The term “router” in this document refers to a router in a generic sense or a Layer 3 switch running the PIM protocol.

PIM Overview

Protocol Independent Multicast (PIM) provides IP multicast forwarding by leveraging static routes or unicast routing tables generated by any unicast routing protocol, such as routing information protocol (RIP), open shortest path first (OSPF), intermediate system to intermediate system (IS-IS), or border gateway protocol (BGP). Independent of the unicast routing protocols running on the device, multicast routing can be implemented as long as the corresponding multicast routing entries are created through unicast routes. PIM uses the reverse path forwarding (RPF) mechanism to implement multicast forwarding. When a multicast packet arrives on an interface of the device, it is subject to an RPF check. If the RPF check succeeds, the device creates the corresponding routing entry and forwards the packet; if the RPF check fails, the device discards the packet.

Based on the routing mechanism, PIM falls into two modes:

- Protocol Independent Multicast-Dense Mode (PIM-DM), and
- Protocol Independent Multicast-Sparse Mode (PIM-SM).



To facilitate description, a network comprising PIM-capable routers is referred to as a “PIM domain” in this document.

Introduction to PIM-DM

PIM-DM is a type of dense mode multicast protocol. It uses the “push mode” for multicast forwarding, and is suitable for small-sized networks with densely distributed multicast members.

The basic implementation of PIM-DM is as follows:

- PIM-DM assumes that at least one multicast group member exists on each subnet of a network, and therefore multicast data is flooded to all nodes on the network. Then, branches without multicast forwarding are pruned from the forwarding tree, leaving only those branches that contain receivers. This “flood and prune” process takes place periodically, that is, pruned branches resume multicast forwarding when the pruned state times out and then data is re-flooded down these branches, and then are pruned again.
- When a new receiver on a previously pruned branch joins a multicast group, to reduce the join latency, PIM-DM uses a graft mechanism to resume data forwarding to that branch.

Generally speaking, the multicast forwarding path is a source tree, namely a forwarding tree with the multicast source as its “root” and multicast group members as its “leaves”. Because the source tree is the shortest path from the multicast source to the receivers, it is also called shortest path tree (SPT).

How PIM-DM Works

The working mechanism of PIM-DM is summarized as follows:

- Neighbor discovery
- SPT building
- Graft
- Assert

Neighbor discovery

In a PIM domain, a PIM router discovers PIM neighbors, maintains PIM neighboring relationships with other routers, and builds and maintains SPTs by periodically multicasting hello messages to all other PIM routers (224.0.0.13).



Every activated interface on a router sends hello messages periodically, and thus learns the PIM neighboring information pertinent to the interface.

SPT establishment

The process of building an SPT is the process of “flood and prune”.

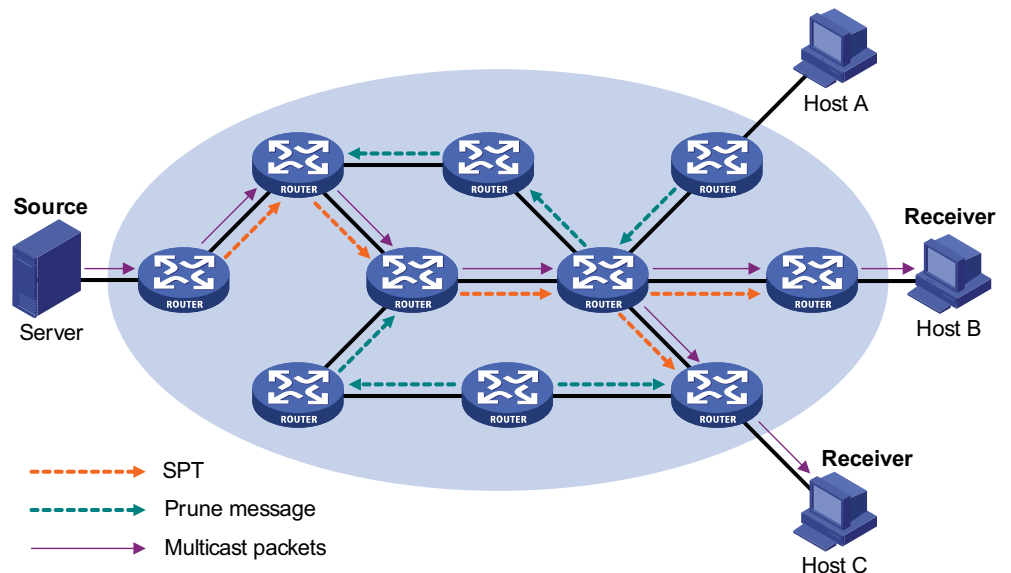
- 1 In a PIM-DM domain, when a multicast source *S* sends multicast data to a multicast group *G*, the multicast packet is first flooded throughout the domain: The router first performs RPF check on the multicast packet. If the packet passes the RPF check, the router creates an (*S*, *G*) entry and forwards the data to all downstream nodes in the network. In the flooding process, an (*S*, *G*) entry is created on all the routers in the PIM-DM domain.
- 2 Then, nodes without receivers downstream are pruned: A router having no receivers downstream sends a prune message to the upstream node to “tell” the upstream node to delete the corresponding interface from the outgoing interface list in the (*S*, *G*) entry and stop forwarding subsequent packets addressed to that multicast group down to this node.



- *An (*S*, *G*) entry contains the multicast source address *S*, multicast group address *G*, outgoing interface list, and incoming interface.*
- *For a given multicast stream, the interface that receives the multicast stream is referred to as “upstream”, and the interfaces that forward the multicast stream are referred to as “downstream”.*

A prune process is first initiated by a leaf router. As shown in Figure 187, a router without any receiver attached to it (the router connected with Host A, for example) sends a prune message, and this prune process goes on until only necessary branches are left in the PIM-DM domain. These branches constitute the SPT.

Figure 187 SPT establishment



The “flood and prune” process takes place periodically. A pruned state timeout mechanism is provided. A pruned branch restarts multicast forwarding when the pruned state times out and then is pruned again when it no longer has any multicast receiver.



Pruning has a similar implementation in PIM-SM.

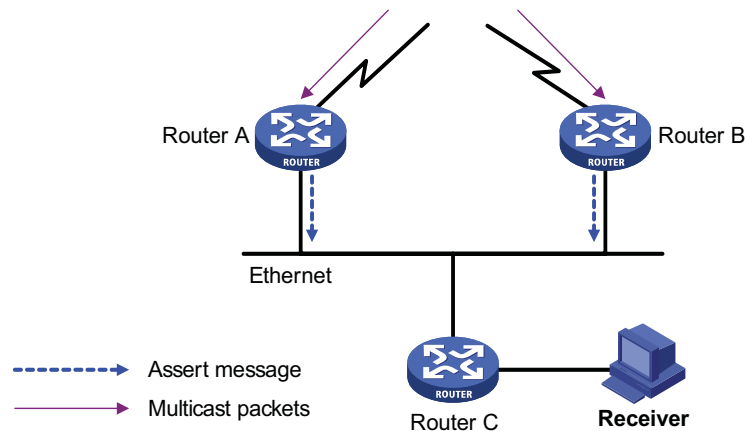
Graft

When a host attached to a pruned node joins a multicast group, to reduce the join latency, PIM-DM uses a graft mechanism to resume data forwarding to that branch. The process is as follows:

- 1 The node that needs to receive multicast data sends a graft message hop by hop toward the source, as a request to join the SPT again.
- 2 Upon receiving this graft message, the upstream node puts the interface on which the graft was received into the forwarding state and responds with a graft-ack message to the graft sender.
- 3 If the node that sent a graft message does not receive a graft-ack message from its upstream node, it will keep sending graft messages at a configurable interval until it receives an acknowledgment from its upstream node.

Assert

If multiple multicast routers exist on a multi-access subnet, duplicate packets may flow to the same subnet. To shut off duplicate flows, the assert mechanism is used for election of a single multicast forwarder on a multi-access network.

Figure 188 Assert mechanism

As shown in Figure 188, after Router A and Router B receive an (S, G) packet from the upstream node, they both forward the packet to the local subnet. As a result, the downstream node Router C receives two identical multicast packets, and both Router A and Router B, on their own local interface, receive a duplicate packet forwarded by the other. Upon detecting this condition, both routers send an assert message to all PIM routers (224.0.0.13) through the interface on which the packet was received. The assert message contains the following information: the multicast source address (S), the multicast group address (G), and the preference and metric of the unicast route to the source. By comparing these parameters, either Router A or Router B becomes the unique forwarder of the subsequent (S, G) packets on the multi-access subnet. The comparison process is as follows:

- 1 The router with a higher unicast route preference to the source wins;
- 2 If both routers have the same unicast route preference to the source, the router with a smaller metric to the source wins;
- 3 If there is a tie in route metric to the source, the router with a higher IP address of the local interface wins.

Introduction to PIM-SM

PIM-DM uses the “flood and prune” principle to build SPTs for multicast data distribution. Although an SPT has the shortest path, it is built with a low efficiency. Therefore the PIM-DM mode is not suitable for large- and medium-sized networks.

PIM-SM is a type of sparse mode multicast protocol. It uses the “pull mode” for multicast forwarding, and is suitable for large- and medium-sized networks with sparsely and widely distributed multicast group members.

The basic implementation of PIM-SM is as follows:

- PIM-SM assumes that no hosts need to receive multicast data. In the PIM-SM mode, routers must specifically request a particular multicast stream before the data is forwarded to them. The core task for PIM-SM to implement multicast forwarding is to build and maintain rendezvous point trees (RPTs). An RPT is rooted at a router in the PIM domain as the common node, or rendezvous point (RP), through which the multicast data travels along the RPT and reaches the receivers.

- When a receiver is interested in the multicast data addressed to a specific multicast group, the router connected to this receiver sends a join message to the RP corresponding to that multicast group. The path along which the message goes hop by hop to the RP forms a branch of the RPT.
- When a multicast source sends a multicast packet to a multicast group, the router directly connected with the multicast source first registers the multicast source with the RP by sending a register message to the RP by unicast. The arrival of this message at the RP triggers the establishment of an SPT. Then, the multicast source sends subsequent multicast packets along the SPT to the RP. Upon reaching the RP, the multicast packet is duplicated and delivered to the receivers along the RPT.



Multicast traffic is duplicated only where the distribution tree branches, and this process automatically repeats until the multicast traffic reaches the receivers.

How PIM-SM Works

The working mechanism of PIM-SM is summarized as follows:

- Neighbor discovery
- DR election
- RP discovery
- RPT building
- Multicast source registration
- Switchover from RPT to SPT
- Assert

Neighbor discovery

PIM-SM uses exactly the same neighbor discovery mechanism as PIM-DM does. Refer to “Neighbor discovery” on page 630.

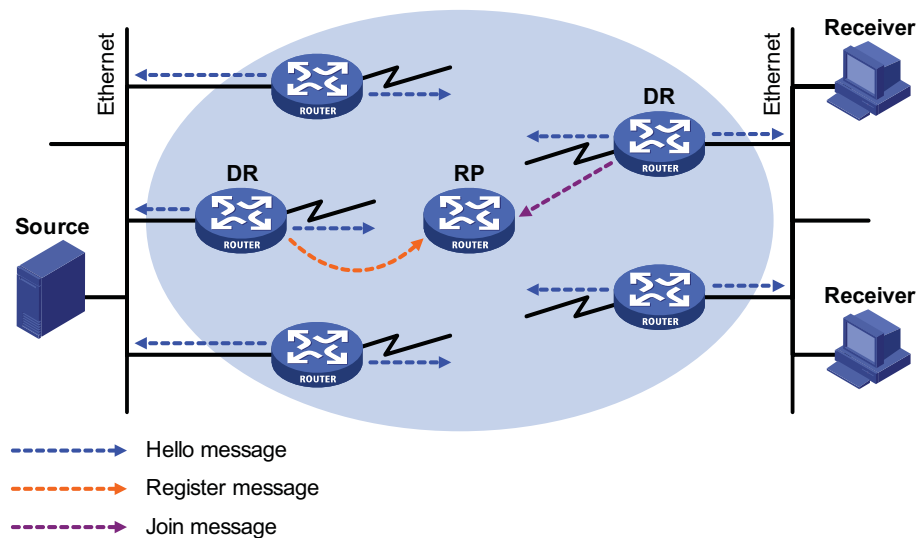
DR election

PIM-SM also uses hello messages to elect a designated router (DR) for a multi-access network. The elected DR will be the only multicast forwarder on this multi-access network.

A DR must be elected in a multi-access network, no matter this network connects to multicast sources or to receivers. The DR at the receiver side sends join messages to the RP; the DR at the multicast source side sends register messages to the RP.



- *A DR is elected on a multi-access subnet by means of comparison of the priorities and IP addresses carried in hello messages. An elected DR is substantially meaningful to PIM-SM. PIM-DM itself does not require a DR. However, if IGMPv1 runs on any multi-access network in a PIM-DM domain, a DR must be elected to act as the IGMPv1 querier on that multi-access network.*
- *IGMP must be enabled on a device that acts as a DR before receivers attached to this device can join multicast groups through this DR.*

Figure 189 DR election

As shown in Figure 189, the DR election process is as follows:

- 1 Routers on the multi-access network send hello messages to one another. The hello messages contain the router priority for DR election. The router with the highest DR priority will become the DR.
- 2 In the case of a tie in the router priority, or if any router in the network does not support carrying the DR-election priority in hello messages, the router with the highest IP address will win the DR election.

When the DR fails, a timeout in receiving hello message triggers a new DR election process among the other routers.

RP discovery

The RP is the core of a PIM-SM domain. For a small-sized, simple network, one RP is enough for forwarding information throughout the network, and the position of the RP can be statically specified on each router in the PIM-SM domain. In most cases, however, a PIM-SM network covers a wide area and a huge amount of multicast traffic needs to be forwarded through the RP. To lessen the RP burden and optimize the topological structure of the RPT, each multicast group should have its own RP. Therefore, a bootstrap mechanism is needed for dynamic RP election. For this purpose, a bootstrap router (BSR) should be configured.

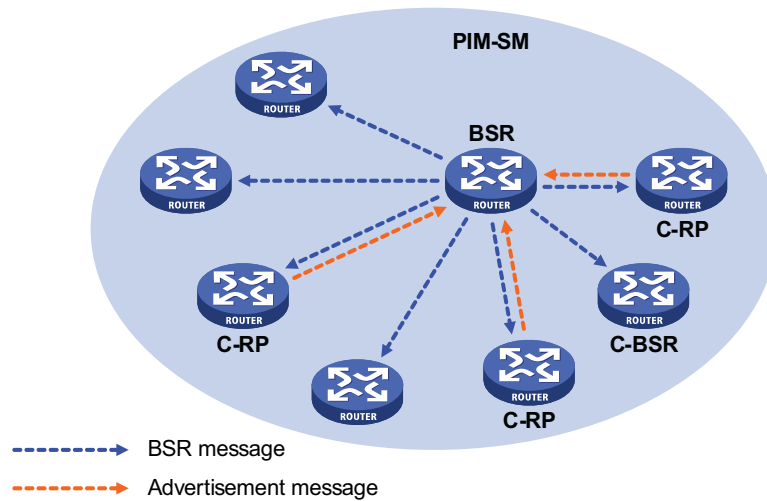
As the administrative core of a PIM-SM domain, the BSR collects advertisement messages (C-RP-Adv messages) from candidate-RPs (C-RPs) and chooses the appropriate C-RP information for each multicast group to form an RP-set, which is a database of mappings between multicast groups and RPs. The BSR then floods the RP-set to the entire PIM-SM domain. Based on the information in these RP-sets, all routers (including the DRs) in the network can calculate the location of the corresponding RPs.

A PIM-SM domain (or an administratively scoped region) can have only one BSR, but can have multiple candidate-BSRs (C-BSRs). Once the BSR fails, a new BSR is automatically elected from the C-BSRs through the bootstrap mechanism to avoid service interruption. Similarly, multiple C-RPs can be configured in a PIM-SM

domain, and the position of the RP corresponding to each multicast group is calculated through the BSR mechanism.

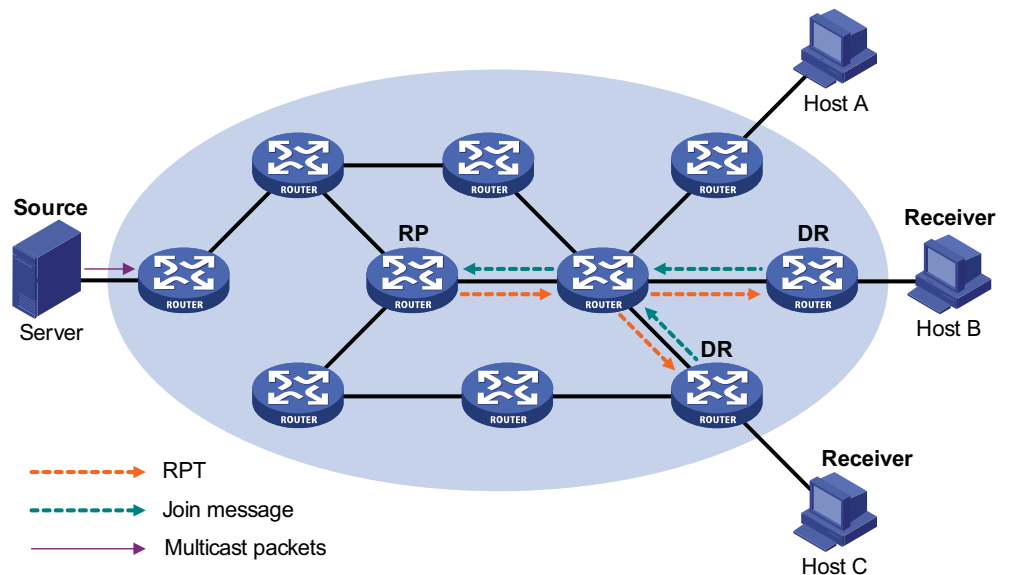
Figure 190 shows the positions of C-RPs and the BSR in the network.

Figure 190 BSR and C-RPs



RPT establishment

Figure 191 RPT establishment in a PIM-SM domain



As shown in Figure 191, the process of building an RPT is as follows:

- 1 When a receiver joins a multicast group G, it uses an IGMP message to inform the directly connected DR.
- 2 Upon getting the receiver information, the DR sends a join message, which is hop by hop forwarded to the RP corresponding to the multicast group.
- 3 The routers along the path from the DR to the RP form an RPT branch. Each router on this branch generates a (*, G) entry in its forwarding table. The * means any multicast source. The RP is the root, while the DRs are the leaves, of the RPT.

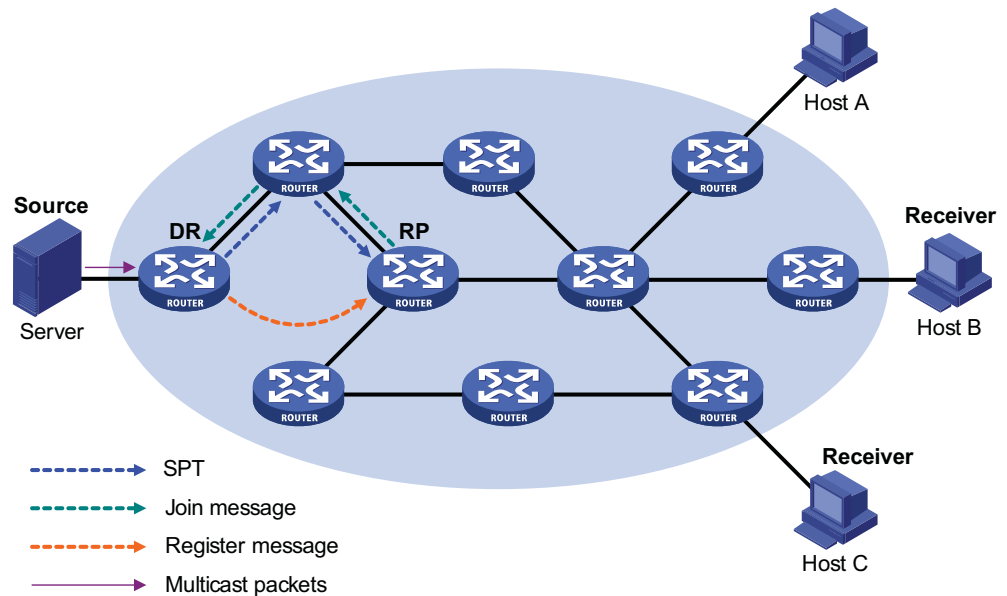
The multicast data addressed to the multicast group G flows through the RP, reaches the corresponding DR along the established RPT, and finally is delivered to the receiver.

When a receiver is no longer interested in the multicast data addressed to a multicast group G, the directly connected DR sends a prune message, which goes hop by hop along the RPT to the RP. Upon receiving the prune message, the upstream node deletes its link with this downstream node from the outgoing interface list and checks whether it itself has receivers for that multicast group. If not, the router continues to forward the prune message to its upstream router.

Multicast source registration

The purpose of multicast source registration is to inform the RP about the existence of the multicast source.

Figure 192 Multicast registration



As shown in Figure 192, the multicast source registers with the RP as follows:

- 1 When the multicast source S sends the first multicast packet to a multicast group G, the DR directly connected with the multicast source, upon receiving the multicast packet, encapsulates the packet in a PIM register message, and sends the message to the corresponding RP by unicast.
- 2 When the RP receives the register message, it extracts the multicast packet from the register message and forwards the multicast packet down the RPT, and sends an (S, G) join message hop by hop toward the multicast source. Thus, the routers along the path from the RP to the multicast source constitute an SPT branch. Each router on this branch generates an (S, G) entry in its forwarding table. The multicast source is the root, while the RP is the leaf, of the SPT.
- 3 The subsequent multicast data from the multicast source travels along the established SPT to the RP, and then the RP forwards the data along the RPT to the receivers. When the multicast traffic arrives at the RP along the SPT, the RP sends a register-stop message to the source-side DR by unicast to stop the source registration process.

Switchover from RPT to SPT

Initially, multicast traffic flows along an RPT from the RP to the receivers. Because the RPT is not necessarily the tree that has the shortest path, upon receiving the first multicast packet along the RPT (by default), or when detecting that the multicast traffic rate reaches a configurable threshold (if so configured), the receiver-side DR initiates an RPT-to-SPT switchover process, as follows:

- 1 First, the receiver-side DR sends an (S, G) join message hop by hop to the multicast source. When the join message reaches the source-side DR, all the routers on the path have installed the (S, G) entry in their forwarding table, and thus an SPT branch is established.
- 2 Subsequently, the receiver-side DR sends a prune message hop by hop to the RP. Upon receiving this prune message, the RP forwards it toward the multicast source, thus to implement RPT-to-SPT switchover.

After the RPT-to-SPT switchover, multicast data can be directly sent from the source to the receivers. PIM-SM builds SPTs through RPT-to-SPT switchover more economically than PIM-DM does through the “flood and prune” mechanism.

Assert

PIM-SM uses exactly the same assert mechanism as PIM-DM does. Refer to “Assert” on page 631.

Introduction to BSR Admin-scope Regions in PIM-SM

Division of PIM-SM domains

Typically, a PIM-SM domain contains only one BSR, which is responsible for advertising RP-set information within the entire PIM-SM domain. The information for all multicast groups is forwarded within the network scope administered by the BSR.

To implement refined management and group-specific services, a PIM-SM domain can be divided into one global scope zone and multiple BSR administratively scoped regions (BSR admin-scope regions).

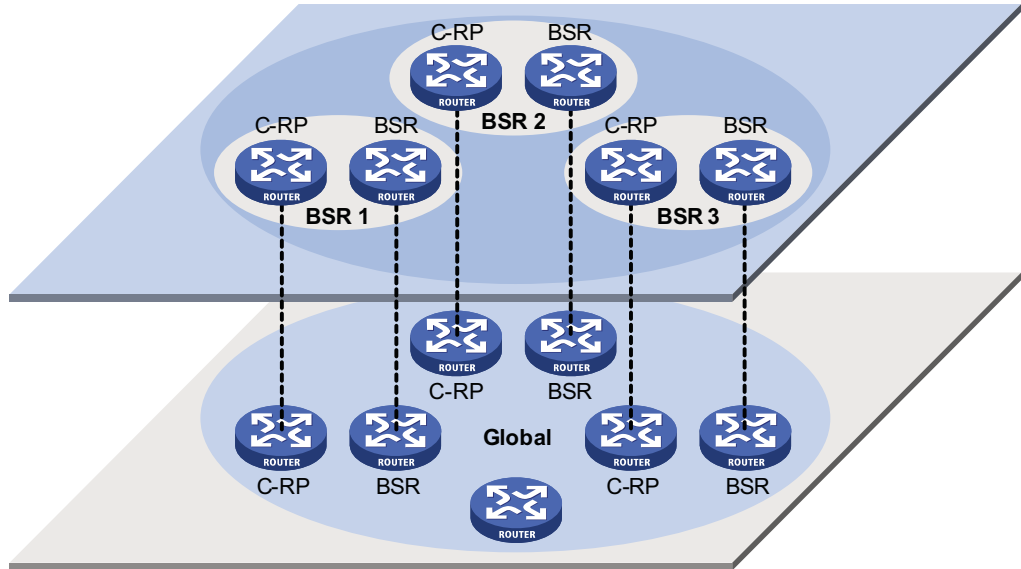
Specific to particular multicast groups, the BSR administrative scoping mechanism effectively lessens the management workload of a single-BSR domain and provides group-specific services.

Relationship between BSR admin-scope regions and the global scope zone

A better understanding of the global scope zone and BSR admin-scope regions should be based on two aspects: geographical space and group address range.

- 1 Geographical space
BSR admin-scope regions are logical regions specific to particular multicast groups, and each BSR admin-scope region must be geographically independent of every other one, as shown in Figure 193.

Figure 193 Relationship between BSR admin-scope regions and the global scope zone in geographic space

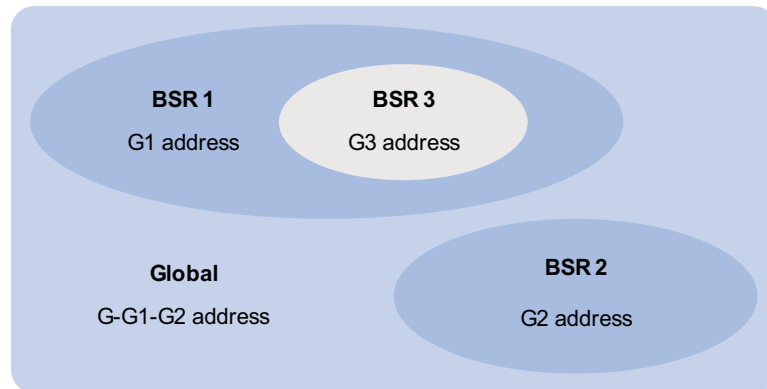


BSR admin-scope regions are geographically separated from one another. Namely, a router must not serve different BSR admin-scope regions. In other words, different BSR admin-scope regions contain different routers, whereas the global scope zone covers all routers in the PIM-SM domain.

2 In terms of multicast group address ranges

Each BSR admin-scope region serves specific multicast groups. Usually, these addresses have no intersections; however, they may overlap one another.

Figure 194 Relationship between BSR admin-scope regions and the global scope zone in group address ranges



In Figure 194, the group address ranges of admin-scope regions BSR1 and BSR2 have no intersection, whereas the group address range of BSR3 is a subset of the address range of BSR1. The group address range of the global scope zone covers all the group addresses other than those of all the BSR admin-scope regions. That is, the group address range of the global scope zone is G-G1-G2. In other words, there is a supplementary relationship between the global scope zone and all the BSR admin-scope regions in terms of group address ranges.

Relationships between BSR admin-scope regions and the global scope zone are as follows:

- The global scope zone and each BSR admin-scope region have their own C-RPs and BSR. These devices are effective only in their respective admin-scope regions. Namely, the BSR election and RP election are implemented independently within each admin-scope region.
- Each BSR admin-scope region has its own boundary. The multicast information (such as C-RP-Adv messages and BSR bootstrap messages) can be transmitted only within the domain.
- Likewise, the multicast information in the global scope zone cannot enter any BSR admin-scope region.
- In terms of multicast information propagation, BSR admin-scope regions are independent of one another and each BSR admin-scope region is independent of the global scope zone, and no overlapping is allowed between any two BSR admin-scope regions.

SSM Model Implementation in PIM

The source-specific multicast (SSM) model and the any-source multicast (ASM) model are two opposite models. Presently, the ASM model includes the PIM-DM and PIM-SM modes. The SSM model can be implemented by leveraging part of the PIM-SM technique.

The SSM model provides a solution for source-specific multicast. It maintains the relationships between hosts and routers through IGMPv3.

In actual application, part of the PIM-SM technique is adopted to implement the SSM model. In the SSM model, receivers know exactly where a multicast source is located by means of advertisements, consultancy, and so on. Therefore, no RP is needed, no RPT is required, there is no source registration process, and there is no need of using the multicast source discovery protocol (MSDP) for discovering sources in other PIM domains.

Compared with the ASM model, the SSM model only needs the support of IGMPv3 and some subsets of PIM-SM. The operation mechanism of PIM-SSM can be summarized as follows:

- Neighbor discovery
- DR election
- SPT building

Neighbor discovery

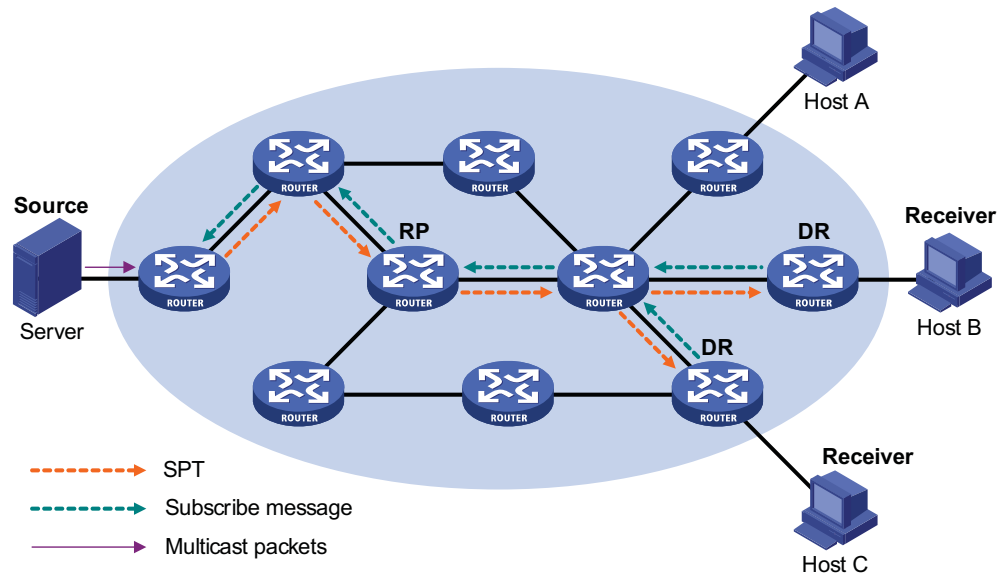
PIM-SSM uses the same neighbor discovery mechanism as in PIM-DM and PIM-SM. Refer to “Neighbor discovery” on page 630.

DR election

PIM-SSM uses the same DR election mechanism as in PIM-SM. Refer to “DR election” on page 633.

Construction of SPT

Whether to build an RPT for PIM-SM or an SPT for PIM-SSM depends on whether the multicast group the receiver is to join falls in the SSM group range (SSM group range reserved by IANA is 232.0.0.0/8).

Figure 195 SPT establishment in PIM-SSM

As shown in Figure 195, Host B and Host C are multicast information receivers. They send IGMPv3 report messages denoted as (Include S, G) to the respective DRs to express their interest in the information of the specific multicast source S. If they need information from other sources than S, they send an (Exclude S, G) report. No matter what the description is, the position of multicast source S is explicitly specified for receivers.

The DR that has received the report first checks whether the group address in this message falls in the SSM group range:

- If so, the DR sends a subscribe message for channel subscription hop by hop toward the multicast source S. An (Include S, G) or (Exclude S, G) entry is created on all routers on the path from the DR to the source. Thus, an SPT is built in the network, with the source S as its root and receivers as its leaves. This SPT is the transmission channel in PIM-SSM.
- If not, the PIM-SM process is followed: the DR needs to send a (*, G) join message to the RP, and a multicast source registration process is needed.



In PIM-SSM, the “channel” concept is used to refer to a multicast group, and the “channel subscription” concept is used to refer to a join message.

Protocols and Standards

PIM-related specifications are as follows:

- RFC 2362: Protocol Independent Multicast-sparse Mode (PIM-SM): Protocol Specification
- RFC 3973: Protocol Independent Multicast-Dense Mode (PIM-DM): Protocol Specification(Revised)
- draft-ietf-pim-sm-v2-new-06: Protocol Independent Multicast-Sparse Mode (PIM-SM)
- draft-ietf-pim-dm-new-v2-02: Protocol Independent Multicast-Dense Mode (PIM-DM)

- draft-ietf-pim-v2-dm-03: Protocol Independent Multicast Version 2 Dense Mode Specification
- draft-ietf-pim-sm-bsr-03: Bootstrap Router (BSR) Mechanism for PIM Sparse Mode
- draft-ietf-ssm-arch-02: Source-Specific Multicast for IP
- draft-ietf-ssm-overview-04: An Overview of Source-Specific Multicast (SSM)

Configuring PIM-DM

PIM-DM Configuration Task List

Complete these tasks to configure PIM-DM:

Task	Remarks
"Enabling PIM-DM" on page 641	Required
"Enabling State Refresh" on page 642	Optional
"Configuring State Refresh Parameters" on page 642	Optional
"Configuring PIM-DM Graft Retry Period" on page 643	Optional
"Configuring PIM Common Information" on page 653	Optional

Configuration Prerequisites

Before configuring PIM-DM, complete the following task:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.

Before configuring PIM-DM, prepare the following data:

- The interval between state refresh messages
- Minimum time to wait before receiving a new refresh message
- TTL value of state refresh messages
- Graft retry period

Enabling PIM-DM

With PIM-DM enabled, a router sends hello messages periodically to discover PIM neighbors and processes messages from PIM neighbors. When deploying a PIM-DM domain, you are recommended to enable PIM-DM on all interfaces of non-border routers (border routers are PIM-enabled routers located on the boundary of BSR admin-scope regions).

Follow these steps to enable PIM-DM:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable IP multicast routing	multicast routing-enable	Required Disable by default
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable PIM-DM	pim dm	Required Disabled by default

**CAUTION:**

- All the interfaces of the same router must work in the same PIM mode.
- PIM-DM cannot be used for multicast groups in the SSM group range.

Enabling State Refresh

An interface without the state refresh capability cannot forward state refresh messages.

Follow these steps to enable the state refresh capability:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable state refresh	pim state-refresh-capable	Optional Enabled by default

Configuring State Refresh Parameters

To avoid the resource-consuming reflooding of unwanted traffic caused by timeout of pruned interfaces, the router directly connected with the multicast source periodically sends an (S, G) state refresh message, which is forwarded hop by hop along the initial multicast flooding path of the PIM-DM domain, to refresh the prune timer state of all the routers on the path.

A router may receive multiple state refresh messages within a short time, of which some may be duplicated messages. To keep a router from receiving such duplicated messages, you can configure the time the router must wait before receiving the next state refresh message. If a new state refresh message is received within the waiting time, the router will discard it; if this timer times out, the router will accept a new state refresh message, refresh its own PIM state, and reset the waiting timer.

The TTL value of a state refresh message decrements by 1 whenever it passes a router before it is forwarded to the downstream node until the TTL value comes down to 0. In a small network, a state refresh message may cycle in the network. To effectively control the propagation scope of state refresh messages, you need to configure an appropriate TTL value based on the network size.

Follow these steps to configure state refresh parameters:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PIM view	pim	-
Configure the interval between state refresh messages	state-refresh-interval <i>interval</i>	Optional 60 seconds by default
Configure the time to wait before receiving a new state refresh message	state-refresh-rate-limit <i>interval</i>	Optional 30 seconds by default
Configure the TTL value of state refresh messages	state-refresh-ttl <i>tvl-value</i>	Optional 255 by default

Configuring PIM-DM Graft Retry Period

In PIM-DM, graft is the only type of message that uses the acknowledgment mechanism. In a PIM-DM domain, if a router does not receive a graft-ack message from the upstream router within the specified time after it sends a graft message, the router keeps sending new graft messages at a configurable interval, namely graft retry period, until it receives a graft-ack from the upstream router.

Follow these steps to configure graft retry period:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure graft retry period	pim timer graft-retry <i>interval</i>	Optional 3 seconds by default



For the configuration of other timers in PIM-DM, refer to “Configuring PIM Common Timers” on page 656.

Configuring PIM-SM



A device can serve as a C-RP and a C-BSR at the same time.

PIM-SM Configuration Task List

Complete these tasks to configure PIM-SM:

Task	Remarks
“Configuring PIM-SM” on page 643	Required
“Configuring a BSR” on page 644	Optional
“Performing basic C-BSR configuration” on page 645	Optional
“Configuring a global-scope C-BSR” on page 646	Optional
“Configuring an admin-scope C-BSR” on page 646	Optional
“Configuring a BSR admin-scope region boundary” on page 647	Optional
“Configuring global C-BSR parameters” on page 647	Optional
“Configuring an RP” on page 648	Optional
“Configuring a static RP” on page 648	Optional
“Configuring a C-RP” on page 648	Optional
“Enabling auto-RP” on page 649	Optional
“Configuring C-RP timers” on page 649	Optional
“Configuring PIM-SM Register Messages” on page 650	Optional
“Disabling RPT-to-SPT Switchover” on page 651	Optional
“Configuring PIM Common Information” on page 653	Optional

Configuration Prerequisites

Before configuring PIM-SM, complete the following task:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.

Before configuring PIM-SM, prepare the following data:

- An ACL rule defining a legal BSR address range
- Hash mask length for RP selection calculation
- C-BSR priority
- Bootstrap interval
- Bootstrap timeout time
- An ACL rule defining a legal C-RP address range and the range of multicast groups to be served
- C-RP-Adv interval
- C-RP timeout time
- The IP address of a static RP
- An ACL rule for register message filtering
- Register suppression timeout time
- Probe time
- ACL rules and ACL order for disabling RPT-to-SPT switchover

Enabling PIM-SM

With PIM-SM enabled, a router sends hello messages periodically to discover PIM neighbors and processes messages from PIM neighbors. When deploying a PIM-SM domain, you are recommended to enable PIM-SM on all interfaces of non-border routers (border routers are PIM-enabled routers located on the boundary of BSR admin-scope regions).

Follow these steps to enable PIM-SM:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable IP multicast routing	multicast routing-enable	Required Disable by default
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable PIM-SM	pim sm	Required Disabled by default



CAUTION: All the interfaces of the same router must work in the same PIM mode.

Configuring a BSR



The BSR is dynamically elected from a number of C-BSRs. Because it is unpredictable which router will finally win a BSR election, the commands introduced in this section must be configured on all C-BSRs.

About the Hash mask length and C-BSR priority for RP selection calculation

- You can configure these parameters at three levels: global configuration level, global scope level, and BSR admin-scope level.
- By default, the global scope parameters and BSR admin-scope parameters are those configured at the global configuration level.
- Parameters configured at the global scope level or BSR admin-scope level have higher priority than those configured at the global configuration level.

Performing basic C-BSR configuration

A PIM-SM domain can have only one BSR, but must have at least one C-BSR. Any router can be configured as a C-BSR. Elected from C-BSRs, a BSR is responsible for collecting and advertising RP information in the PIM-SM.

C-BSRs should be configured on routers in the backbone network. When configuring a router as a C-BSR, make sure that router is PIM-SM enabled. The BSR election process is as follows:

- Initially, every C-BSR assumes itself to be the BSR of this PIM-SM domain, and uses its interface IP address as the BSR address to send bootstrap messages.
- When a C-BSR receives the bootstrap message of another C-BSR, it first compares its own priority with the other C-BSR's priority carried in the message. The C-BSR with a higher priority wins. If there is a tie in the priority, the C-BSR with a higher IP address wins. The loser uses the winner's BSR address to replace its own BSR address and no longer assumes itself to be the BSR, while the winner keeps its own BSR address and continues assuming itself to be the BSR.

Configuring a legal range of BSR addresses enables filtering of BSR messages based on the address range, thus to prevent malicious hosts from initiating attacks by disguising themselves as legitimate BSRs. To protect legitimate BSRs from being maliciously replaced, preventive measures are taken specific to the following two situations:

- 1 Some malicious hosts intend to fool routers by forging BSR messages and change the RP mapping relationship. Such attacks often occur on border routers. Because a BSR is inside the network whereas hosts are outside the network, you can protect a BSR against attacks from external hosts by enabling border routers to perform neighbor check and RPF check on BSR messages and discard unwanted messages.
- 2 When a router in the network is controlled by an attacker or when an illegal router is present in the network, the attacker can configure such a router to be a C-BSR and make it win BSR election so as to gain the right of advertising RP information in the network. After being configured as a C-BSR, a router automatically floods the network with BSR messages. As a BSR message has a TTL value of 1, the whole network will not be affected as long as the neighbor router discards these BSR messages. Therefore, if a legal BSR address range is configured on all routers in the entire network, all routers will discard BSR messages from out of the legal address range, and thus this kind of attacks can be prevented.

The above-mentioned preventive measures can partially protect the security of BSRs in a network. However, if a legal BSR is controlled by an attacker, the above-mentioned problem will also occur.

Follow these steps to complete basic C-BSR configuration:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PIM view	pim	-
Configure an interface as a C-BSR	c-bsr <i>interface-type</i> <i>interface-number</i> [<i>hash-length</i> [<i>priority</i>]]	Required No C-BSR is configured by default
Configure a legal BSR address range	bsr-policy <i>acl-number</i>	Optional No restrictions on BSR address range by default



Since a large amount of information needs to be exchanged between a BSR and the other devices in the PIM-SM domain, a relatively large bandwidth should be provided between the C-BSR and the other devices in the PIM-SM domain.

Configuring a global-scope C-BSR

Follow these steps to configure a global-scope C-BSR:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PIM view	pim	-
Configure a global-scope C-BSR	c-bsr global [hash-length <i>hash-length</i> priority <i>priority</i>] *	Required No global-scope C-BSRs by default

Configuring an admin-scope C-BSR

By default, a PIM-SM domain has only one BSR. The entire network should be managed by this BSR. To manage your network more effectively and specifically, you can divide a PIM-SM domain into multiple BSR admin-scope regions, with each BSR admin-scope region having one BSR, which serves specific multicast groups.

Specific to particular multicast groups, the BSR administrative scoping mechanism effectively lessens the management workload of a single-BSR domain and provides group-specific services.

In a network divided into BSR admin-scope regions, BSRs are elected from multitudinous C-BSRs to serve different multicast groups. The C-RPs in a BSR admin-scope region send C-RP-Adv messages to only the corresponding BSR. The BSR summarizes the advertisement messages into an RP-set and advertises it to all the routers in the BSR admin-scope region. All the routers use the same algorithm to get the RP addresses corresponding to specific multicast groups.

Follow these steps to configure an admin-scope C-BSR:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PIM view	pim	-

To do...	Use the command...	Remarks
Enable BSR administrative scoping	c-bsr admin-scope	Required Disabled by default
Configure an admin-scope C-BSR	c-bsr group <i>group-address</i> { <i>mask</i> <i>mask-length</i> } [hash-length <i>hash-length</i> priority <i>priority</i>] *	Optional No admin-scope BSRs by default

Configuring a BSR admin-scope region boundary

A BSR has its specific service scope. A number of BSR boundary interfaces divide a network into different BSR admin-scope regions. Bootstrap messages cannot cross the admin-scope region boundary, while other types of PIM messages can.

Follow these steps to configure a BSR admin-scope region boundary:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure a BSR admin-scope region boundary	pim bsr-boundary	Required No BSR admin-scope region boundary by default

Configuring global C-BSR parameters

The BSR election winner advertises its own IP address and RP-set information throughout the region it serves through bootstrap messages. The BSR floods bootstrap messages throughout the network periodically. Any C-BSR that receives a bootstrap message maintains the BSR state for a configurable period of time (BSR state timeout), during which no BSR election takes place. When the BSR state times out, a new BSR election process will be triggered among the C-BSRs.

Follow these steps to configure global C-BSR parameters:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PIM view	pim	-
Configure the Hash mask length for RP selection calculation	c-bsr hash-length <i>hash-length</i>	Optional 30 by default
Configure the C-BSR priority	c-bsr priority <i>priority</i>	Optional 0 by default
Configure the bootstrap interval	c-bsr interval <i>interval</i>	Optional For the system default, see "Note" below.
Configure the bootstrap timeout time	c-bsr holdtime <i>interval</i>	Optional For the system default, see "Note" below.



About the bootstrap timeout time

- By default, the bootstrap timeout time is determined by this formula: $\text{Bootstrap timeout} = \text{Bootstrap interval} \times 2 + 10$. The default bootstrap interval is 60 seconds, so the default bootstrap timeout = $60 \times 2 + 10 = 130$ (seconds).
- If this parameter is manually configured, the system will use the configured value.

About the bootstrap interval

- By default, the bootstrap interval is determined by this formula: $\text{Bootstrap interval} = (\text{Bootstrap timeout} - 10) / 2$. The default bootstrap timeout is 130 seconds, so the default bootstrap interval = $(130 - 10) / 2 = 60$ (seconds).
- If this parameter is manually configured, the system will use the configured value.



CAUTION: In configuration, make sure that the bootstrap interval is smaller than the bootstrap timeout time.

Configuring an RP

An RP can be manually configured or dynamically elected through the BSR mechanism. For a large PIM network, static RP configuration is a tedious job. Generally, static RP configuration is just a backup means for the dynamic RP election mechanism to enhance the robustness and operation manageability of a multicast network.

Configuring a static RP

If there is only one dynamic RP in a network, manually configuring a static RP can avoid communication interruption due to single-point failures and avoid frequent message exchange between C-RPs and the BSR. To enable a static RP to work normally, you must perform this configuration on all the devices in the PIM-SM domain and specify the same RP address.

Follow these steps to configure a static RP

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PIM view	pim	-
Configure a static RP	static-rp <i>rp-address</i> [<i>acl-number</i>] [preferred]	Optional No static RP by default

Configuring a C-RP

In a PIM-SM domain, you can configure routers that intend to become the RP as C-RPs. The BSR collects the C-RP information by receiving the C-RP-Adv messages from C-RPs or auto-RP announcements from other routers and organizes the information into an RP-set, which is flooded throughout the entire network. Then, the other routers in the network calculate the mappings between specific group ranges and the corresponding RPs based on the RP-set. We recommend that you configure C-RPs on backbone routers.

To guard against C-RP spoofing, you need to configure a legal C-RP address range and the range of multicast groups to be served on the BSR. In addition, because

every C-BSR has a chance to become the BSR, you need to configure the same filtering policy on all C-BSRs.

Follow these steps to configure a C-RP:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PIM view	pim	-
Configure an interface to be a C-RP	c-rp <i>interface-type interface-number</i> [group-policy <i>acl-number</i> priority <i>priority</i> holdtime <i>hold-interval</i> advertisement-interval <i>adv-interval</i>] *	Optional No C-RPs are configured by default
Configure a legal C-RP address range and the range of multicast groups to be served	crp-policy <i>acl-number</i>	Optional No restrictions by default



- *When configuring a C-RP, ensure a relatively large bandwidth between this C-RP and the other devices in the PIM-SM domain.*
- *An RP can serve multiple multicast groups or all multicast groups. Only one RP can forward multicast traffic for a multicast group at a moment.*

Enabling auto-RP

Auto-RP announcement and discovery messages are respectively addressed to the multicast group addresses 224.0.1.39 and 224.0.1.40. With auto-RP enabled on a device, the device can receive these two types of messages and record the RP information carried in such messages.

Follow these steps to enable auto-RP:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PIM view	pim	-
Enable auto-RP	auto-rp enable	Optional Disabled by default

Configuring C-RP timers

To enable the BSR to distribute the RP-set information within the PIM-SM domain, C-RPs must periodically send C-RP-Adv messages to the BSR. The BSR learns the RP-set information from the received messages, and encapsulates its own IP address together with the RP-set information in its bootstrap messages. The BSR then floods the bootstrap messages to all PIM routers (224.0.0.13) in the network.

Each C-RP encapsulates a timeout value in its C-RP-Adv message. Upon receiving this message, the BSR obtains this timeout value and starts a C-RP timeout timer. If the BSR fails to hear a subsequent C-RP-Adv message from the C-RP when the timer times out, the BSR assumes the C-RP to have expired or become unreachable.

Follow these steps to configure C-RP timers:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PIM view	pim	-
Configure the C-RP-Adv interval	c-rp advertisement-interval <i>interval</i>	Optional 60 seconds by default
Configure C-RP timeout time	c-rp holdtime <i>interval</i>	Optional 150 seconds by default



- *The commands introduced in this section are to be configured on C-RPs.*
- *For the configuration of other timers in PIM-SM, refer to “Configuring PIM Common Timers” on page 656.*

Configuring PIM-SM Register Messages

Within a PIM-SM domain, the source-side DR sends register messages to the RP, and these register messages have different multicast source or group addresses. You can configure a filtering rule to filter register messages so that the RP can serve specific multicast groups. If an (S, G) entry is denied by the filtering rule, or the action for this entry is not defined in the filtering rule, the RP will send a register-stop message to the DR to stop the registration process for the multicast data.

In view of information integrity of register messages in the transmission process, you can configure the device to calculate the checksum based on the entire register messages. However, to reduce the workload of encapsulating data in register messages and for the sake of interoperability, this method of checksum calculation is not recommended.

When receivers stop receiving multicast data addressed to a certain multicast group through the RP (that is, the RP stops serving the receivers of a specific multicast group), or when the RP formally starts receiving multicast data from the multicast source, the RP sends a register-stop message to the source-side DR. Upon receiving this message, the DR stops sending register messages encapsulated with multicast data and enters the register suppression state.

In a probe suppression cycle, the DR can send a null register message (a register message without multicast data encapsulated), a certain length of time defined by the probe time before the register suppression timer expires, to the RP to indicate that the multicast source is active. When the register suppression timer expires, the DR starts sending register messages again. A smaller register suppression timeout setting will cause the RP to receive bursting multicast data more frequently, while a larger timeout setting will result in a larger delay for new receivers to join the multicast group they are interested in.

Follow these steps to configure PIM-SM register-related parameters:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PIM view	pim	-

To do...	Use the command...	Remarks
Configure a filtering rule for register messages	register-policy <i>acl-number</i>	Optional No register filtering rule by default
Configure the device to calculate the checksum based on the entire register messages	register-header-checksum	Optional By default, the checksum is calculated based on the header of register messages
Configure the register suppression timeout time	register-suppression-timeout <i>interval</i>	Optional 60 seconds by default
Configure the probe time	probe-interval <i>interval</i>	Optional 5 seconds by default



Typically, you need to configure the above-mentioned parameters on the receiver-side DR and the RP only. Since both the DR and RP are elected, however, you should carry out these configurations on the routers that may win the DR election and on the C-RPs that may win RP elections.

Disabling RPT-to-SPT Switchover

Initially, multicast traffic flows along an RPT to the receivers. By default, the last-hop switch initiates an RPT-to-SPT switchover process when it receives the first multicast packet from the RPT. You can disable RPT-to-SPT switchover through the following configuration.

Follow these steps to disable RPT-to-SPT switchover:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PIM view	pim	-
Disable RPT-to-SPT switchover	spt-switch-threshold infinity [group-policy <i>acl-number</i> [order <i>order-value</i>]]	Optional By default, the device switches to the SPT immediately after it receives the first multicast packet from the RPT.



- The support for the **timer spt-switch** command depends on the specific device model.
- Typically, you need to configure the above-mentioned parameters on the receiver-side DR and the RP only. Since both the DR and RP are elected, however, you should carry out these configurations on the routers that may win the DR election and on the C-RPs that may win RP elections.
- If the multicast source is learned through MSDP, the device will switch to the SPT immediately after it receives the first multicast packet from the RPT, no matter how big the traffic rate threshold is set (this threshold is not configurable on a switch).

Configuring PIM-SSM



The PIM-SSM model needs the support of IGMPv3. Therefore, be sure to enable IGMPv3 on PIM routers with multicast receivers.

PIM-SSM Configuration Task List

Complete these tasks to configure PIM-SSM:

Task	Remarks
"Enabling PIM-SM" on page 652	Required
"Configuring the SSM Group Range" on page 652	Optional
"Configuring PIM Common Information" on page 653	Optional

Configuration Prerequisites

Before configuring PIM-SSM, complete the following task:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.

Before configuring PIM-SSM, prepare the following data:

- The SSM group range

Enabling PIM-SM

The SSM model is implemented based on some subsets of PIM-SM. Therefore, a router is PIM-SSM capable after you enable PIM-SM on it.

When deploying a PIM-SM domain, you are recommended to enable PIM-SM on all interfaces of non-border routers (border routers are PIM-enabled routers located on the boundary of BSR admin-scope regions).

Follow these steps to enable PIM-SM:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable IP multicast routing	multicast routing-enable	Required Disabled by default
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable PIM-SM	pim sm	Required Disabled by default



CAUTION: All the interfaces of the same router must work in the same PIM mode.

Configuring the SSM Group Range

As for whether the information from a multicast source is delivered to the receivers based on the PIM-SSM model or the PIM-SM model, this depends on whether the group address in the (S, G) channel subscribed by the receivers falls in the SSM group range. All PIM-SM-enabled interfaces assume that multicast groups within this address range are using the PIM-SSM model.

Follow these steps to configure an SSM multicast group range:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PIM view	pim	-
Configure the SSM group range	ssm-policy <i>acl-number</i>	Optional 232.0.0.0/8 by default



The commands introduced in this section are to be configured on all routers in the PIM domain.



CAUTION:

- Make sure that the same SSM group range is configured on all routers in the entire domain. Otherwise, multicast information cannot be delivered through the SSM model.
- When a member of a multicast group in the SSM group range sends an IGMPv1 or IGMPv2 report message, the device does not trigger a (*, G) join.

Configuring PIM Common Information



For the configuration tasks described in this section

- Configurations performed in PIM view are effective to all interfaces, while configurations performed in interface view are effective to the current interface only.
- If the same function or parameter is configured in both PIM view and interface view, the configuration performed in interface view is given priority, regardless of the configuration sequence.

PIM Common Information Configuration Task List

Complete these tasks to configure PIM common information:

Task	Remarks
"Configuring a PIM Filter" on page 654	Optional
"Configuring PIM Hello Options" on page 654	Optional
"Configuring PIM Common Timers" on page 656	Optional
"Configuring Join/Prune Message Limits" on page 657	Optional

Configuration Prerequisites

Before configuring PIM common information, complete the following tasks:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Configure PIM-DM, or PIM-SM, or PIM-SSM.

Before configuring PIM common information, prepare the following data:

- An ACL rule as multicast data filter
- Priority for DR election (global value/interface level value)
- PIM neighbor timeout time (global value/interface value)

- Prune delay (global value/interface level value)
- Prune override interval (global value/interface level value)
- Hello interval (global value/interface level value)
- Maximum delay between hello message (interface level value)
- Assert timeout time (global value/interface value)
- Join/prune interval (global value/interface level value)
- Join/prune timeout (global value/interface value)
- Multicast source lifetime
- Maximum size of join/prune messages
- Maximum number of (S, G) entries in a join/prune message

Configuring a PIM Filter

No matter in a PIM-DM domain or a PIM-SM domain, routers can check passing-by multicast data based on the configured filtering rules and determine whether to continue forwarding the multicast data. In other words, PIM routers can act as multicast data filters. These filters can help implement traffic control on one hand, and control the information available to receivers downstream to enhance data security on the other hand.

Follow these steps to configure a PIM filter:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PIM view	pim	-
Configure a multicast group filter	source-policy <i>acl-number</i>	Required No multicast data filter by default



- *Generally, a smaller distance from the filter to the multicast source results in a more remarkable filtering effect.*
- *This filter works not only on independent multicast data but also on multicast data encapsulated in register messages.*

Configuring PIM Hello Options

No matter in a PIM-DM domain or a PIM-SM domain, the hello messages sent among routers contain many configurable options, including:

- **DR_Priority** (for PIM-SM only): priority for DR election. The device with the highest priority wins the DR election. You can configure this parameter on all the routers in a multi-access network directly connected to multicast sources or receivers.
- **Holdtime**: the timeout time of PIM neighbor reachability state. When this timer times out, if the router has received no hello message from a neighbor, it assumes that this neighbor has expired or become unreachable. You can configure this parameter on all routers in the PIM domain. If you configure different values for this timer on different neighboring routers, the largest value will take effect.
- **LAN_Prune_Delay**: the delay of prune messages on a multi-access network. This option consists of LAN-delay (namely, prune delay), override-interval, and

neighbor tracking flag bit. You can configure this parameter on all routers in the PIM domain. If different LAN-delay or override-interval values result from the negotiation among all the PIM routers, the largest value will take effect.

The LAN-delay setting will cause the upstream routers to delay processing received prune messages. If the LAN-delay setting is too small, it may cause the upstream router to stop forwarding multicast packets before a downstream router sends a prune override message. Therefore, be cautious when configuring this parameter.

The override-interval sets the length of time a downstream router is allowed to wait before sending a prune override message. When a router receives a prune message from a downstream router, it does not perform the prune action immediately; instead, it maintains the current forwarding state for a period of time defined by LAN-delay. If the downstream router needs to continue receiving multicast data, it must send a prune override message within the prune override interval; otherwise, the upstream route will perform the prune action when the LAN-delay timer times out.

A hello message sent from a PIM router contains a generation ID option. The generation ID is a random value for the interface on which the hello message is sent. Normally, the generation ID of a PIM router does not change unless the status of the router changes (for example, when PIM is just enabled on the interface or the device is restarted). When the router starts or restarts sending hello messages, it generates a new generation ID. If a PIM router finds that the generation ID in a hello message from the upstream router has changed, it assumes that the status of the upstream neighbor is lost or the upstream neighbor has changed. In this case, it triggers a join message for state update.

If you disable join suppression (namely, enable neighbor tracking), the upstream router will explicitly track which downstream routers are joined to it. The join suppression feature should be enabled or disabled on all PIM routers on the same subnet.

Configuring hello options globally

Follow these steps to configure hello options globally:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PIM view	pim	-
Configure the priority for DR election	hello-option dr-priority <i>priority</i>	Optional 1 by default
Configure PIM neighbor timeout time	hello-option holdtime <i>interval</i>	Optional 105 seconds by default
Configure the prune delay time (LAN-delay)	hello-option lan-delay <i>interval</i>	Optional 500 milliseconds by default
Configure the prune override interval	hello-option override-interval <i>interval</i>	Optional 2,500 milliseconds by default
Disable join suppression	hello-option neighbor-tracking	Optional Enabled by default

Configuring hello options on an interface

Follow these steps to configure hello options on an interface:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure the priority for DR election	pim hello-option dr-priority <i>priority</i>	Optional 1 by default
Configure PIM neighbor timeout time	pim hello-option holdtime <i>interval</i>	Optional 105 seconds by default
Configure the prune delay time (LAN-delay)	pim hello-option lan-delay <i>interval</i>	Optional 500 milliseconds by default
Configure the prune override interval	pim hello-option override-interval <i>interval</i>	Optional 2,500 milliseconds by default
Disable join suppression	pim hello-option neighbor-tracking	Optional Enabled by default
Configure the interface to reject hello messages without a generation ID	pim require-genid	Optional By default, hello messages without Generation_ID are accepted

Configuring PIM Common Timers

PIM routers discover PIM neighbors and maintain PIM neighboring relationships with other routers by periodically sending out hello messages.

Upon receiving a hello message, a PIM router waits a random period, which is equal to or smaller than the maximum delay between hello messages, before sending out a hello message. This avoids collisions that occur when multiple PIM routers send hello messages simultaneously.

Any router that has lost assert election will prune its downstream interface and maintain the assert state for a period of time. When the assert state times out, the assert losers will resume multicast forwarding.

A PIM router periodically sends join/prune messages to its upstream for state update. A join/prune message contains the join/prune timeout time. The upstream router sets a join/prune timeout timer for each pruned downstream interface, and resumes the forwarding state of the pruned interface when this timer times out.

When a router fails to receive subsequent multicast data from the multicast source S, the router will not immediately delete the corresponding (S, G) entries; instead, it maintains (S, G) entries for a period of time, namely the multicast source lifetime, before deleting the (S, G) entries.

Configuring PIM common timers globally

Follow these steps to configure PIM common timers globally:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Enter PIM view	pim	-
Configure the hello interval	timer hello <i>interval</i>	Optional 30 seconds by default
Configure assert timeout time	holdtime assert <i>interval</i>	Optional 180 seconds by default
Configure the join/prune interval	timer join-prune <i>interval</i>	Optional 60 seconds by default
Configure the join/prune timeout time	holdtime join-prune <i>interval</i>	Optional 210 seconds by default
Configure the multicast source lifetime	source-lifetime <i>interval</i>	Optional 210 seconds by default

Configuring PIM common timers on an interface

Follow these steps to configure PIM common timers on an interface:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure the hello interval	pim timer hello <i>interval</i>	Optional 30 seconds by default
Configure the maximum delay between hello messages	pim triggered-hello-delay <i>interval</i>	Optional 5 seconds by default
Configure assert timeout time	pim holdtime assert <i>interval</i>	Optional 180 seconds by default
Configure the join/prune interval	pim timer join-prune <i>interval</i>	Optional 60 seconds by default
Configure the join/prune timeout time	pim holdtime join-prune <i>interval</i>	Optional 210 seconds by default



If there are no special networking requirements, we recommend that you use the default settings.

Configuring Join/Prune Message Limits

A larger join/prune message size will result in loss of a larger amount of information when a message is lost; with a reduced join/message size, the loss of a single message will bring relatively minor impact.

By controlling the maximum number of (S, G) entries in a join/prune message, you can effectively reduce the number of (S, G) entries sent per unit of time.

Follow these steps to configure join/prune message limits:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PIM view	pim	-

To do...	Use the command...	Remarks
Configure the maximum size of a join/prune message	jp-pkt-size <i>packet-size</i>	Optional 8,100 bytes by default
Configure the maximum number of (S, G) entries in a join/prune message	jp-queue-size <i>queue-size</i>	Optional 1,020 by default

Displaying and Maintaining PIM

To do...	Use the command...	Remarks
View the BSR information in the PIM-SM domain and locally configured C-RP information in effect	display pim bsr-info	Available in any view
View the information of unicast routes used by PIM	display pim claimed-route [<i>source-address</i>]	Available in any view
View the number of PIM control messages	display pim control-message counters [message-type { probe register register-stop }] [interface <i>interface-type interface-number</i> message-type { assert bsr crp graft graft-ack hello join-prune state-refresh }] *]	Available in any view
View the information about unacknowledged graft messages	display pim grafts	Available in any view
View the PIM information on an interface or all interfaces	display pim interface [<i>interface-type interface-number</i>] [verbose]	Available in any view
View the information of join/prune messages to send	display pim join-prune mode { sm [flags <i>flag-value</i>] ssm } [interface <i>interface-type interface-number</i> neighbor <i>neighbor-address</i>] * [verbose]	Available in any view
View PIM neighboring information	display pim neighbor [interface <i>interface-type interface-number</i> <i>neighbor-address</i>] [verbose] *	Available in any view
View the content of the PIM routing table	display pim routing-table [<i>group-address</i> [mask { <i>mask-length</i> <i>mask</i> }] <i>source-address</i> [mask { <i>mask-length</i> <i>mask</i> }]] incoming-interface [<i>interface-type interface-number</i> register] outgoing-interface { include exclude match } { <i>interface-type interface-number</i> register } mode <i>mode-type</i> flags <i>flag-value</i> fsm] *	Available in any view
View the RP information	display pim rp-info [<i>group-address</i>]	Available in any view
Reset PIM control message counters	reset pim control-message counters [interface <i>interface-type interface-number</i>]	Available in user view

PIM Configuration Examples

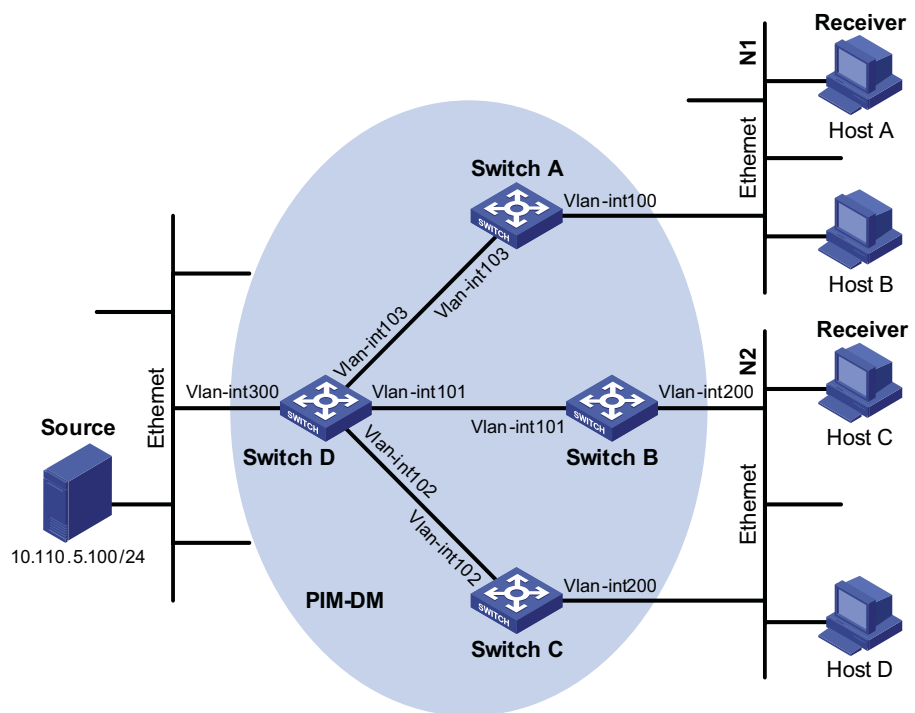
PIM-DM Configuration Example

Network requirements

- Receivers receive VOD information through multicast. The receiver groups of different organizations form stub networks, and one or more receiver hosts exist in each stub network. The entire PIM domain operates in the dense mode.
- Host A and Host C are multicast receivers in two stub networks.
- Switch D connects to the network that comprises the multicast source (Source) through VLAN-interface 300.
- Switch A connects to stub network N1 through VLAN-interface 100, and to Switch D through VLAN-interface 103.
- Switch B and Switch C connect to stub network N2 through their respective VLAN-interface 200, and to Switch D through VLAN-interface 101 and VLAN-interface 102 respectively.
- IGMPv2 is to run between Switch A and N1, and between Switch B/Switch C and N2.

Network diagram

Figure 196 Network diagram for PIM-DM configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	10.110.1.1/24	Switch D	Vlan-int300	10.110.5.1/24
	Vlan-int103	192.168.1.1/24		Vlan-int103	192.168.1.2/24
Switch B	Vlan-int200	10.110.2.1/24		Vlan-int101	192.168.2.2/24
	Vlan-int101	192.168.2.1/24		Vlan-int102	192.168.3.2/24

Switch C	Vlan-int200	10.110.2.2/24
	Vlan-int102	192.168.3.1/24

Configuration procedure

- 1 Configure the interface IP addresses and unicast routing protocol for each switch
Configure the IP address and subnet mask for each interface as per Figure 196. Detailed configuration steps are omitted here.

Configure the OSPF protocol for interoperability among the switches in the PIM-DM domain. Ensure the network-layer interoperability among Switch A, Switch B, Switch C and Switch D in the PIM-DM domain and enable dynamic update of routing information among the switches through a unicast routing protocol. Detailed configuration steps are omitted here.

- 2 Enable IP multicast routing, and enable PIM-DM on each interface

Enable IP multicast routing on Switch A, enable PIM-DM on each interface, and enable IGMPv2 on VLAN-interface 100, which connects Switch A to the stub network.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] igmp enable
[SwitchA-Vlan-interface100] pim dm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 103
[SwitchA-Vlan-interface103] pim dm
[SwitchA-Vlan-interface103] quit
```

The configuration on Switch B and Switch C is similar to that on Switch A.

Enable IP multicast routing on Switch D, and enable PIM-DM on each interface.

```
<SwitchD> system-view
[SwitchD] multicast routing-enable
[SwitchD] interface vlan-interface 300
[SwitchD-Vlan-interface300] pim dm
[SwitchD-Vlan-interface300] quit
[SwitchD] interface vlan-interface 103
[SwitchD-Vlan-interface103] pim dm
[SwitchD-Vlan-interface103] quit
[SwitchD] interface vlan-interface 101
[SwitchD-Vlan-interface101] pim dm
[SwitchD-Vlan-interface101] quit
[SwitchD] interface vlan-interface 102
[SwitchD-Vlan-interface102] pim dm
[SwitchD-Vlan-interface102] quit
```

- 3 Verify the configuration

Use the **display pim interface** command to view the PIM configuration and running status on each interface. For example:

View the PIM configuration information on Switch D.

```
[SwitchD] display pim interface
Interface          NbrCnt HelloInt   DR-Pri   DR-Address
Vlan300            0       30           1        10.110.5.1   (local)
Vlan103            1       30           1        192.168.1.2  (local)
```



```
Vlan101          1      30      1      192.168.2.2    (local)
Vlan102          1      30      1      192.168.3.2    (local)
```

Carry out the **display pim neighbor** command to view the PIM neighboring relationships among the switches. For example:

View the PIM neighboring relationships on Switch D.

```
[SwitchD] display pim neighbor
Total Number of Neighbors = 3
```

Neighbor	Interface	Uptime	Expires	Dr-Priority
192.168.1.1	Vlan103	00:02:22	00:01:27	1
192.168.2.1	Vlan101	00:00:22	00:01:29	3
192.168.3.1	Vlan102	00:00:23	00:01:31	5

Assume that Host A needs to receive the information addressed to a multicast group G (225.1.1.1/24). After multicast source S (10.110.5.100/24) sends multicast packets to the multicast group G, an SPT is established through traffic flooding. Switches on the SPT path (Switch A and Switch D) have their (S, G) entries. Host A registers with Switch A, and a (*, G) entry is generated on Switch A. You can use the **display pim routing-table** command to view the PIM routing table information on each switch. For example:

View the PIM routing table information on Switch A.

```
[SwitchA] display pim routing-table
Total 1 (*, G) entry; 1 (S, G) entry
```

```
(*, 225.1.1.1)
  Protocol: pim-dm, Flag: WC
  UpTime: 00:04:25
  Upstream interface: NULL
    Upstream neighbor: NULL
    RPF prime neighbor: NULL
  Downstream interface(s) information:
    Total number of downstreams: 1
      1: Vlan-interface100
        Protocol: igmp, UpTime: 00:04:25, Expires: never
(10.110.5.100, 225.1.1.1)
  Protocol: pim-dm, Flag: ACT
  UpTime: 00:06:14
  Upstream interface: Vlan-interface103,
    Upstream neighbor: 192.168.1.2
    RPF prime neighbor: 192.168.1.2
  Downstream interface(s) information:
    Total number of downstreams: 1
      1: Vlan-interface100
        Protocol: pim-dm, UpTime: 00:04:25, Expires: never
```

The information on Switch B and Switch C is similar to that on Switch A.

View the PIM routing table information on Switch D.

```
[SwitchD] display pim routing-table
Total 0 (*, G) entry; 1 (S, G) entry
```

```
(10.110.5.100, 225.1.1.1)
  Protocol: pim-dm, Flag: LOC ACT
  UpTime: 00:03:27
  Upstream interface: Vlan-interface300
    Upstream neighbor: NULL
    RPF prime neighbor: NULL
```

```

Downstream interface(s) information:
Total number of downstreams: 3
  1: Vlan-interface103
      Protocol: pim-dm, UpTime: 00:03:27, Expires: never
  2: Vlan-interface101
      Protocol: pim-dm, UpTime: 00:03:27, Expires: never
  3: Vlan-interface102
      Protocol: pim-dm, UpTime: 00:03:27, Expires: never

```

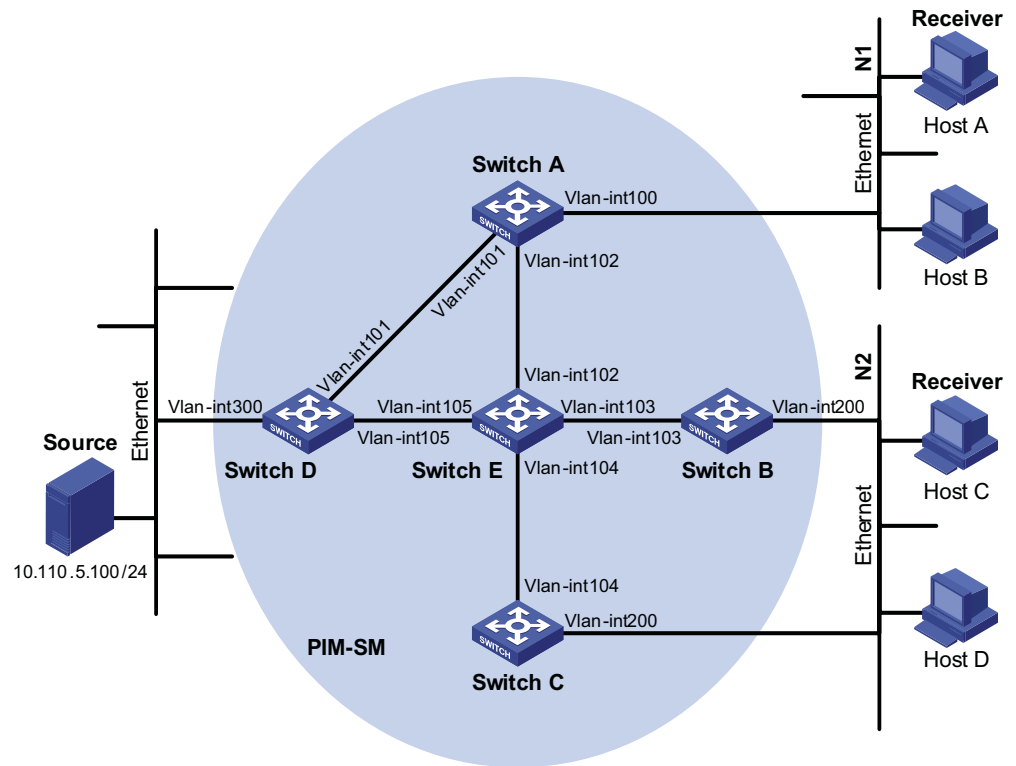
PIM-SM Configuration Example

Network requirements

- Receivers receive VOD information through multicast. The receiver groups of different organizations form stub networks, and one or more receiver hosts exist in each stub network. The entire PIM domain operates in the sparse mode (not divided into different BSR admin-scope regions).
- Host A and Host C are multicast receivers in two stub networks.
- Switch D connects to the network that comprises the multicast source (Source) through VLAN-interface 300.
- Switch A connects to stub network N1 through VLAN-interface 100, and to Switch D and Switch E through VLAN-interface 101 and VLAN-interface 102 respectively.
- Switch B and Switch C connect to stub network N2 through their respective VLAN-interface 200, and to Switch E through VLAN-interface 103 and VLAN-interface 104 respectively.
- Switch E connects to Switch A, Switch B, Switch C and Switch D, and its VLAN-interface 102 interface acts a C-BSR and a C-RP, with the range of multicast groups served by the C-RP being 225.1.1.0/24.
- IGMPv2 is to run between Switch A and N1, and between Switch B/Switch C and N2.

Network diagram

Figure 197 Network diagram for PIM-SM domain configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	10.110.1.1/24	Switch D	Vlan-int300	10.110.5.1/24
	Vlan-int101	192.168.1.1/24		Vlan-int101	192.168.1.2/24
	Vlan-int102	192.168.9.1/24		Vlan-int105	192.168.4.2/24
Switch B	Vlan-int200	10.110.2.1/24	Switch E	Vlan-int104	192.168.3.2/24
	Vlan-int103	192.168.2.1/24		Vlan-int103	192.168.2.2/24
Switch C	Vlan-int200	10.110.2.2/24		Vlan-int102	192.168.9.2/24
	Vlan-int104	192.168.3.1/24	Vlan-int105	192.168.4.1/24	

Configuration procedure

- 1 Configure the interface IP addresses and unicast routing protocol for each switch
Configure the IP address and subnet mask for each interface as per Figure 197. Detailed configuration steps are omitted here.

Configure the OSPF protocol for interoperability among the switches in the PIM-SM domain. Ensure the network-layer interoperability among Switch A, Switch B, Switch C, Switch D and Switch E in the PIM-SM domain and enable dynamic update of routing information among the switches through a unicast routing protocol. Detailed configuration steps are omitted here.

- 2 Enable IP multicast routing, and enable PIM-SM on each interface
Enable IP multicast routing on Switch A, enable PIM-SM on each interface, and enable IGMPv2 on VLAN-interface 100, which connects Switch A to the stub network.

```

<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] igmp enable
[SwitchA-Vlan-interface100] pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim sm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] pim sm
[SwitchA-Vlan-interface102] quit

```

The configuration on Switch B and Switch C is similar to that on Switch A. The configuration on Switch D and Switch E is also similar to that on Switch A except that it is not necessary to enable IGMP on the corresponding interfaces on these two switches.

3 Configure a C-BSR and a C-RP

Configure the service scope of RP advertisements and the positions of the C-BSR and C-RP on Switch E.

```

<SwitchE> system-view
[SwitchE] acl number 2005
[SwitchE-acl-basic-2005] rule permit source 225.1.1.0 0.0.0.255
[SwitchE-acl-basic-2005] quit
[SwitchE] pim
[SwitchE-pim] c-bsr vlan-interface 102
[SwitchE-pim] c-rp vlan-interface 102 group-policy 2005
[SwitchE-pim] quit

```

4 Verify the configuration

Carry out the **display pim interface** command to view the PIM configuration and running status on each interface. For example:

View the PIM configuration information on Switch A.

```

[SwitchA] display pim interface

```

Interface	NbrCnt	HelloInt	DR-Pri	DR-Address	
Vlan100	0	30	1	10.110.1.1	(local)
Vlan101	1	30	1	192.168.1.2	
Vlan102	1	30	1	192.168.9.2	

To view the BSR election information and the locally configured C-RP information in effect on a switch, use the **display pim bsr-info** command. For example:

View the BSR information and the locally configured C-RP information in effect on Switch A.

```

[SwitchA] display pim bsr-info
Elected BSR Address: 192.168.9.2
  Priority: 0
  Hash mask length: 30
  State: Accept Preferred
  Scope: Not scoped
  Uptime: 01:40:40
  Next BSR message scheduled at: 00:01:42

```

View the BSR information and the locally configured C-RP information in effect on Switch E.

```
[SwitchE] display pim bsr-info
  Elected BSR Address: 192.168.9.2
    Priority: 0
    Hash mask length: 30
    State: Elected
    Scope: Not scoped
    Uptime: 00:00:18
    Next BSR message scheduled at: 00:01:52
  Candidate BSR Address: 192.168.9.2
    Priority: 0
    Hash mask length: 30
    State: Pending
    Scope: Not scoped

Candidate RP: 192.168.9.2(Vlan-interface102)
  Priority: 0
  HoldTime: 150
  Advertisement Interval: 60
  Next advertisement scheduled at: 00:00:48
```

To view the RP information discovered on a switch, use the **display pim rp-info** command. For example:

View the RP information on Switch A.

```
[SwitchA] display pim rp-info
  Vpn-instance: public net
PIM-SM BSR RP information:
  Group/MaskLen: 225.1.1.0/24
    RP: 192.168.9.2
    Priority: 0
    HoldTime: 150
    Uptime: 00:51:45
    Expires: 00:02:22
```

Assume that Host A needs to receive information addressed to the multicast group G (225.1.1.1/24). An RPT will be built between Switch A and Switch E. When the multicast source S (10.110.5.100/24) registers with the RP, an SPT will be built between Switch D and Switch E. Upon receiving multicast data, Switch A immediately switches from the RPT to the SPT. Switches on the RPT path (Switch A and Switch E) have a (*, G) entry, while switches on the SPT path (Switch A and Switch D) have an (S, G) entry. You can use the **display pim routing-table** command to view the PIM routing table information on the switches. For example:

View the PIM routing table information on Switch A.

```
[SwitchA] display pim routing-table
Total 1 (*, G) entry; 1 (S, G) entry
(*, 225.1.1.1)
  RP: 192.168.9.2
  Protocol: pim-sm, Flag: WC
  UpTime: 00:13:46
  Upstream interface: Vlan-interface102,
    Upstream neighbor: 192.168.9.2
    RPF prime neighbor: 192.168.9.2
  Downstream interface(s) information:
    Total number of downstreams: 1
      1: Vlan-interface100
        Protocol: igmp, UpTime: 00:13:46, Expires:00:03:06
(10.110.5.100, 225.1.1.1)
  RP: 192.168.9.2
  Protocol: pim-sm, Flag: SPT ACT
```

```

UpTime: 00:00:42
Upstream interface: Vlan-interface101,
  Upstream neighbor: 192.168.9.2
  RPF prime neighbor: 192.168.9.2
Downstream interface(s) information:
  Total number of downstreams: 1
    1: Vlan-interface100
      Protocol: pim-sm, UpTime: 00:00:42, Expires:00:03:06

```

The information on Switch B and Switch C is similar to that on Switch A.

View the PIM routing table information on Switch D.

```

[SwitchD] display pim routing-table
Total 0 (*, G) entry; 1 (S, G) entry
(10.110.5.100, 225.1.1.1)
  RP: 192.168.9.2
  Protocol: pim-sm, Flag: SPT ACT
  UpTime: 00:00:42
  Upstream interface: Vlan-interface300
    Upstream neighbor: NULL
    RPF prime neighbor: NULL
  Downstream interface(s) information:
    Total number of downstreams: 1
      1: Vlan-interface105
        Protocol: pim-sm, UpTime: 00:00:42, Expires:00:02:06

```

View the PIM routing table information on Switch E.

```

[SwitchE] display pim routing-table
Total 1 (*, G) entry; 0 (S, G) entry

(*, 225.1.1.1)
  RP: 192.168.9.2 (local)
  Protocol: pim-sm, Flag: WC
  UpTime: 00:13:16
  Upstream interface: Register
    Upstream neighbor: 192.168.4.2
    RPF prime neighbor: 192.168.4.2
  Downstream interface(s) information:
    Total number of downstreams: 1
      1: Vlan-interface102
        Protocol: pim-sm, UpTime: 00:13:16, Expires: 00:03:22

```

PIM-SSM Configuration Example

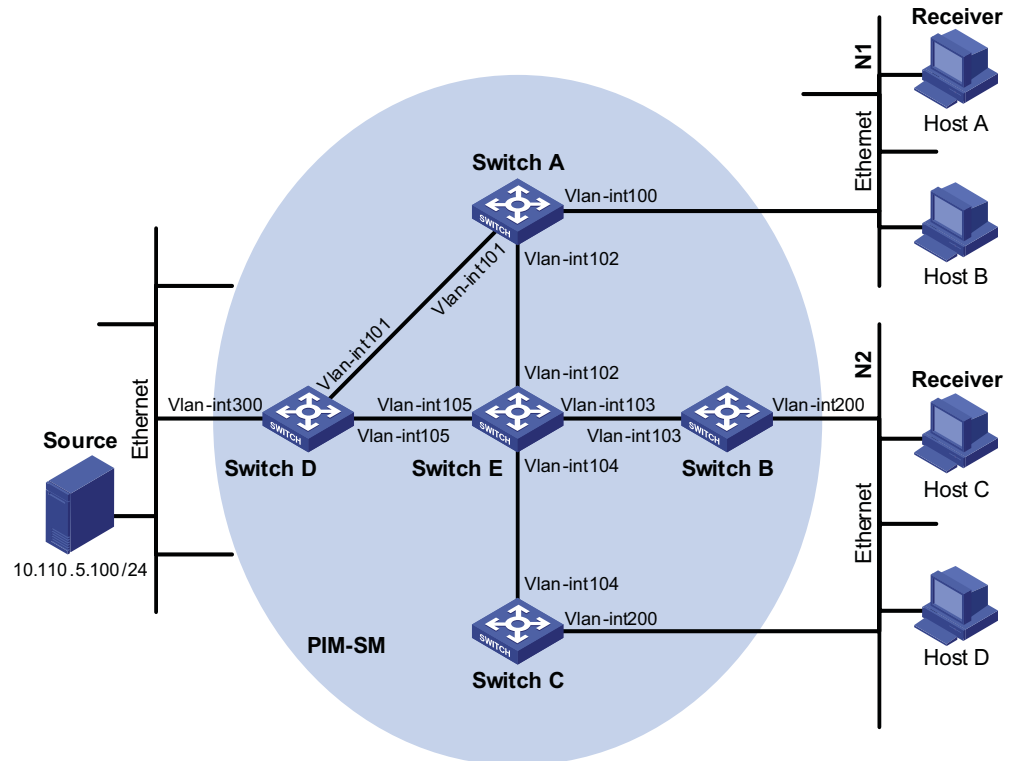
Network requirements

- Receivers receive VOD information through multicast. The receiver groups of different organizations form stub networks, and one or more receiver hosts exist in each stub network. The entire PIM domain operates in the SSM mode.
- Host A and Host C are multicast receivers in two stub networks.
- Switch D connects to the network that comprises the multicast source (Source) through VLAN-interface 300.
- Switch A connects to stub network N1 through VLAN-interface 100, and to Switch D and Switch E through VLAN-interface 101 and VLAN-interface 102 respectively.
- Switch B and Switch C connect to stub network N2 through their respective VLAN-interface 200, and to Switch E through VLAN-interface 103 and VLAN-interface 104 respectively.
- Switch E connects to Switch A, Switch B, Switch C and Switch D.
- The SSM group range is 232.1.1.0/24.

- IGMPv3 is to run between Switch A and N1, and between Switch B/Switch C and N2.

Network diagram

Figure 198 Network diagram for PIM-SSM configuration



Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int100	10.110.1.1/24	Switch D	Vlan-int300	10.110.5.1/24
	Vlan-int101	192.168.1.1/24		Vlan-int101	192.168.1.2/24
	Vlan-int102	192.168.9.1/24		Vlan-int105	192.168.4.2/24
Switch B	Vlan-int200	10.110.2.1/24	Switch E	Vlan-int104	192.168.3.2/24
	Vlan-int103	192.168.2.1/24		Vlan-int103	192.168.2.2/24
Switch C	Vlan-int200	10.110.2.2/24		Vlan-int102	192.168.9.2/24
	Vlan-int104	192.168.3.1/24	Vlan-int105	192.168.4.1/24	

Configuration procedure

- 1 Configure the interface IP addresses and unicast routing protocol for each switch

Configure the IP address and subnet mask for each interface as per Figure 198. Detailed configuration steps are omitted here.

Configure the OSPF protocol for interoperation among the switches in the PIM-SM domain. Ensure the network-layer interoperation among Switch A, Switch B, Switch C, Switch D and Switch E in the PIM-SM domain and enable dynamic update of routing information among the switches through a unicast routing protocol. Detailed configuration steps are omitted here.

- 2 Enable IP multicast routing, and enable PIM-SM on each interface

Enable IP multicast routing on Switch A, enable PIM-SM on each interface, and enable IGMPv3 on VLAN-interface 100, which connects Switch A to the stub network.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] igmp enable
[SwitchA-Vlan-interface100] igmp version 3
[SwitchA-Vlan-interface100] pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 101
[SwitchA-Vlan-interface101] pim sm
[SwitchA-Vlan-interface101] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] pim sm
[SwitchA-Vlan-interface102] quit
```

The configuration on Switch B and Switch C is similar to that on Switch A. The configuration on Switch D and Switch E is also similar to that on Switch A except that it is not necessary to enable IGMP on the corresponding interfaces on these two switches.

3 Configure the SSM group range

Configure the SSM group range to be 232.1.1.0/24 on Switch A.

```
[SwitchA] acl number 2000
[SwitchA-acl-basic-2000] rule permit source 232.1.1.0 0.0.0.255
[SwitchA-acl-basic-2000] quit
[SwitchA] pim
[SwitchA-pim] ssm-policy 2000
[SwitchA-pim] quit
```

The configuration on Switch B, Switch C, Switch D and Switch E is similar to that on Switch A.

4 Verify the configuration

Carry out the **display pim interface** command to view the PIM configuration and running status on each interface. For example:

View the PIM configuration information on Switch A.

```
[SwitchA] display pim interface
Interface           NbrCnt HelloInt   DR-Pri   DR-Address
Vlan100             0       30          1        10.110.1.1  (
local)
Vlan101             1       30          1        192.168.1.2
Vlan102             1       30          1        192.168.9.2
```

Assume that Host A needs to receive the information a specific multicast source S (10.110.5.100/24) sends to multicast group G (232.1.1.1/24). Switch A builds an SPT toward the multicast source. Switches on the SPT path (Switch A and Switch D) have generated an (S, G) entry, while Switch E, which is not on the SPT path, does not have multicast routing entries. You can use the **display pim routing-table** command to view the PIM routing table information on each switch. For example:

View the PIM routing table information on Switch A.

```
[SwitchA] display pim routing-table
Total 0 (*, G) entry; 1 (S, G) entry
```



```
(10.110.5.100, 232.1.1.1)
  Protocol: pim-ssm, Flag:
  UpTime: 00:13:25
  Upstream interface: Vlan-interface101
    Upstream neighbor: 192.168.1.2
    RPF prime neighbor: 192.168.1.2
  Downstream interface(s) information:
    Total number of downstreams: 1
    1: Vlan-interface100
      Protocol: igmp, UpTime: 00:13:25, Expires: -
```

The information on Switch B and Switch C is similar to that on Switch A.

View the PIM routing table information on Switch D.

```
[SwitchD] display pim routing-table
Total 0 (*, G) entry; 1 (S, G) entry
(10.110.5.100, 232.1.1.1)
  Protocol: pim-ssm, Flag:LOC
  UpTime: 00:12:05
  Upstream interface: Vlan-interface300
    Upstream neighbor: NULL
    RPF prime neighbor: NULL
  Downstream interface(s) information:
    Total number of downstreams: 1
    1: Vlan-interface105
      Protocol: pim-ssm, UpTime: 00:12:05, Expires: 00:03:25
```

Troubleshooting PIM Configuration

Failure of Building a Multicast Distribution Tree Correctly

Symptom

None of the routers in the network (including routers directly connected with multicast sources and receivers) has multicast forwarding entries. That is, a multicast distribution tree cannot be built correctly and clients cannot receive multicast data.

Analysis

- When PIM-DM runs on the entire network, multicast data is flooded from the first hop router connected with the multicast source to the last hop router connected with the clients along the SPT. When the multicast data is flooded to a router, no matter which router is, it creates (S, G) entries only if it has a route to the multicast source. If the router does not have a route to the multicast source, or if PIM-DM is not enabled on the router's RPF interface to the multicast source, the router cannot create (S, G) entries.
- When PIM-SM runs on the entire network, and when a router is to join the SPT, the router creates (S, G) entries only if it has a route to the multicast source. If the router does not have a route to the multicast source, or if PIM-DM is not enabled on the router's RPF interface to the multicast source, the router cannot create (S, G) entries.
- When a multicast router receives a multicast packet, it searches the existing unicast routing table for the optimal route to the RPF check object. The outgoing interface of this route will act as the RPF interface and the next hop will be taken as the RPF neighbor. The RPF interface completely relies on the

existing unicast route, and is independent of PIM. The RPF interface must be PIM-enabled, and the RPF neighbor must also be a PIM neighbor. If PIM is not enabled on the router where the RPF interface or the RPF neighbor resides, the establishment of a multicast distribution tree will surely fail, causing abnormal multicast forwarding.

- Because a hello message does not carry the PIM mode information, a router running PIM is unable to know what PIM mode its PIM neighbor is running. If different PIM modes are enabled on the RPF interface and on the corresponding interface of the RPF neighbor router, the establishment of a multicast distribution tree will surely fail, causing abnormal multicast forwarding.
- The same PIM mode must run on the entire network. Otherwise, the establishment of a multicast distribution tree will surely fail, causing abnormal multicast forwarding.

Solution

- 1 Check unicast routes. Use the **display ip routing-table** command to check whether a unicast route exists from the receiver host to the multicast source.
- 2 Check that PIM is enabled on the interfaces, especially on the RPF interface. Use the **display pim interface** command to view the PIM information on each interface. If PIM is not enabled on the interface, use the **pim dm** or **pim sm** command to enable PIM-DM or PIM-SM.
- 3 Check that the RPF neighbor is a PIM neighbor. Use the **display pim neighbor** command to view the PIM neighbor information.
- 4 Check that PIM and IGMP are enabled on the interfaces directly connecting to the multicast source and to the receivers.
- 5 Check that the same PIM mode is enabled on related interfaces. Use the **display pim interface verbose** command to check whether the same PIM mode is enabled on the RPF interface and the corresponding interface of the RPF neighbor router.
- 6 Check that the same PIM mode is enabled on all the routers in the entire network. Make sure that the same PIM mode is enabled on all the routers: PIM-SM on all routers, or PIM-DM on all routers. In the case of PIM-SM, also check that the BSR and RP configurations are correct.

Multicast Data Abnormally Terminated on an Intermediate Router

Symptom

An intermediate router can receive multicast data successfully, but the data cannot reach the last hop router. An interface on the intermediate router receives data but no corresponding (S, G) entry is created in the PIM routing table.

Analysis

- If a multicast forwarding boundary has been configured through the **multicast boundary** command, any multicast packet will be kept from crossing the boundary, and therefore no routing entry can be created in the PIM routing table.
- In addition, the **source-policy** command is used to filter received multicast packets. If the multicast data fails to pass the ACL rule defined in this command, PIM cannot create the route entry, either.

Solution

- 1 Check the multicast forwarding boundary configuration. Use the **display current-configuration** command to check the multicast forwarding boundary settings. Use the **multicast boundary** command to change the multicast forwarding boundary settings.
- 2 Check the multicast filter configuration. Use the **display current-configuration** command to check the multicast filter configuration. Change the ACL rule defined in the **source-policy** command so that the source/group address of the multicast data can pass ACL filtering.

RPs Unable to Join SPT in PIM-SM**Symptom**

An RPT cannot be established correctly, or the RPs cannot join the SPT to the multicast source.

Analysis

- As the core of a PIM-SM domain, the RPs serve specific multicast groups. Multiple RPs can coexist in a network. Make sure that the RP information on all routers is exactly the same, and a specific group is mapped to the same RP. Otherwise, multicast forwarding will fail.
- If the static RP mechanism is used, the same static RP command must be executed on all the routers in the entire network. Otherwise, multicast forwarding will fail.

Solution

- 1 Check that a route is available to the RP. Carry out the **display ip routing-table** command to check whether a route is available on each router to the RP.
- 2 Check the dynamic RP information. Use the **display pim rp-info** command to check whether the RP information is consistent on all routers.
- 3 Check the configuration of static RPs. Use the **display pim rp-info** command to check whether the same static RP address has been configured on all the routers in the entire network.

No Unicast Route Between BSR and C-RPs in PIM-SM**Symptom**

C-RPs cannot unicast advertise messages to the BSR. The BSR does not advertise bootstrap messages containing C-RP information and has no unicast route to any C-RP. An RPT cannot be established correctly, or the DR cannot perform source register with the RP.

Analysis

- The C-RPs periodically send C-RP-Adv messages to the BSR by unicast. If a C-RP has no unicast route to the BSR, the BSR cannot receive C-RP-Adv messages from that C-RP and the bootstrap message of the BSR will not contain the information of that C-RP.
- In addition, if the BSR does not have a unicast router to a C-RP, it will discard the C-RP-Adv messages from that C-RP, and therefore the bootstrap messages of the BSR will not contain the information of that C-RP.

- The RP is the core of a PIM-SM domain. Make sure that the RP information on all routers is exactly the same, a specific group G is mapped to the same RP, and unicast routes are available to the RP.

Solution

- 1 Check whether routes to C-RPs, the RP and the BSR are available. Carry out the **display ip routing-table** command to check whether routes are available on each router to the RP and the BSR, and whether a route is available between the RP and the BSR. Make sure that each C-RP has a unicast route to the BSR, the BSR has a unicast route to each C-RP, and all the routers in the entire network have a unicast route to the RP.
- 2 Check the RP and BSR information. PIM-SM needs the support of the RP and BSR. Use the **display pim bsr-info** command to check whether the BSR information is available on each router, and then use the **display pim rp-info** command to check whether the RP information is correct.
- 3 View PIM neighboring relationships. Use the **display pim neighbor** command to check whether the normal PIM neighboring relationships have been established among the routers

When configuring MSDP, go to these sections for information you are interested in:

- “MSDP Overview” on page 673
- “MSDP Configuration Task List” on page 679 * MERGEFORMAT
- “Displaying and Maintaining MSDP” on page 685
- “MSDP Configuration Examples” on page 685
- “Troubleshooting MSDP” on page 697



The term “router” in this document refers to a router in a generic sense or a Layer 3 switch running the MSDP protocol.

MSDP Overview

Introduction to MSDP

Multicast source discovery protocol (MSDP) is an inter-domain multicast solution developed to address the interconnection of protocol independent multicast sparse mode (PIM-SM) domains. It is used to discover multicast source information in other PIM-SM domains.

In the basic PIM-SM mode, a multicast source registers only with the RP in the local PIM-SM domain, and the multicast source information of a domain is isolated from that of another domain. As a result, the RP is aware of the source information only within the local domain and a multicast distribution tree is built only within the local domain to deliver multicast data from a local multicast source to local receivers. If there is a mechanism that allows RPs of different PIM-SM domains to share their multicast source information, the local RP will be able to join multicast sources in other domains and multicast data can be transmitted among different domains.

MSDP achieves this objective. By establishing MSDP peer relationships among RPs of different PIM-SM domains, source active (SA) messages can be forwarded among domains and the multicast source information can be shared.



CAUTION:

- *MSDP is applicable only if the intra-domain multicast protocol is PIM-SM.*
- *MSDP is meaningful only for the any-source multicast (ASM) model.*

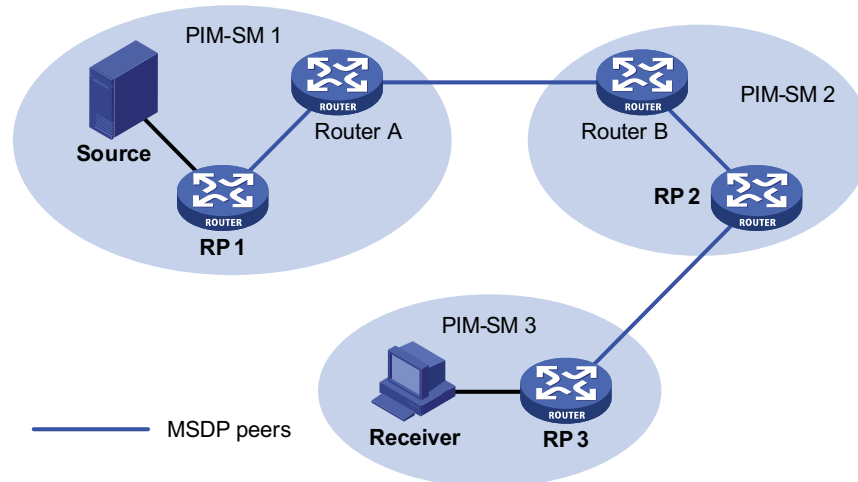
How MSDP Works

MSDP peers

With one or more pairs of MSDP peers configured in the network, an MSDP interconnection map is formed, where the RPs of different PIM-SM domains are

interconnected in series. Relayed by these MSDP peers, an SA message sent by an RP can be delivered to all other RPs.

Figure 199 Where MSDP peers are in the network



As shown in Figure 199, an MSDP peer can be created on any PIM-SM router. MSDP peers created on PIM-SM routers that assume different roles function differently.

- 1 MSDP peers on RPs
- 2 Source-side MSDP peer: the MSDP peer nearest to the multicast source (Source), typically the source-side RP, like RP 1. The source-side RP creates SA messages and sends the messages to its remote MSDP peer to notify the MSDP peer of the locally registered multicast source information. A source-side MSDP must be created on the source-side RP; otherwise it will not be able to advertise the multicast source information out of the PIM-SM domain.
- 3 Receiver-side MSDP peer: the MSDP peer nearest to the receivers, typically the receiver-side RP, like RP 3. Upon receiving an SA message, the receiver-side MSDP peer resolves the multicast source information carried in the message and joins the SPT rooted at the source across the PIM-SM domain. When multicast data from the multicast source arrives, the receiver-side MSDP peer forwards the data to the receivers along the RPT.
- 4 Intermediate MSDP peer: an MSDP peer with multicast remote MSDP peers, like RP 2. An intermediate MSDP peer forwards SA messages received from one remote MSDP peer to other remote MSDP peers, functioning as a relay of multicast source information.
- 5 MSDP peers created on common PIM-SM routers (other than RPs)

Router A and Router B are MSDP peers on common multicast routers. Such MSDP peers just forward received SA messages.

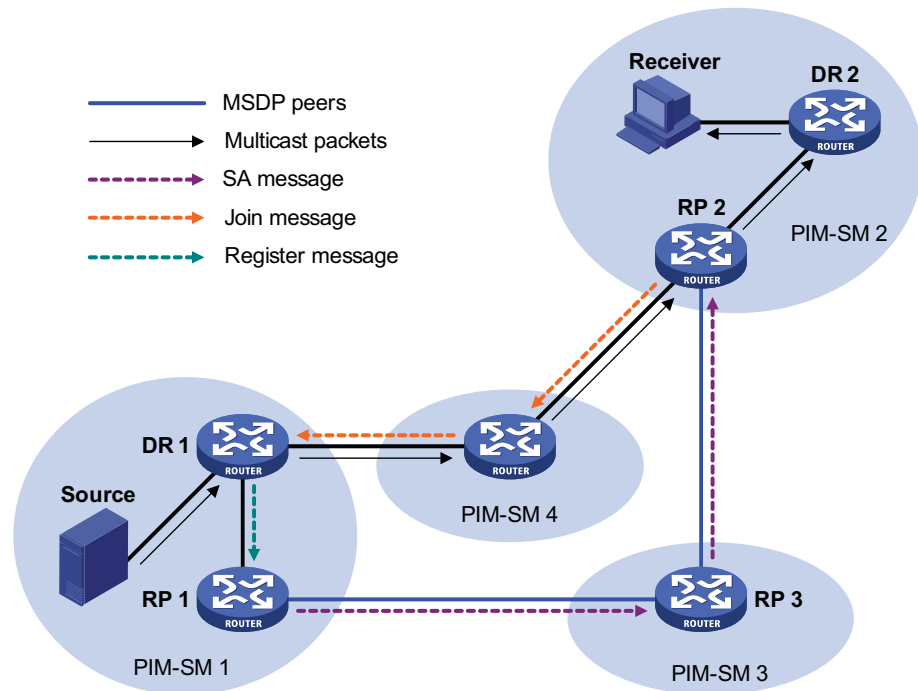


An RP is dynamically elected from C-RPs. To enhance network robustness, a PIM-SM network typically has more than one C-RP. As the RP election result is unpredictable, MSDP peering relationships should be built among all C-RPs so that the winner C-RP is always on the “MSDP interconnection map”, while loser C-RPs will assume the role of common PIM-SM routers on the “MSDP interconnection map”.

Implementing inter-domain multicast delivery by leveraging MSDP peers

As shown in Figure 200, an active source (Source) exists in the domain PIM-SM 1, and RP 1 has learned the existence of Source through multicast source registration. If RPs in PIM-SM 2 and PIM-SM 3 also wish to know the specific location of Source so that receiver hosts can receive multicast traffic originated from it, MSDP peering relationships should be established between RP 1 and RP 3 and between RP 3 and RP 2 respectively.

Figure 200 MSDP peering relationships



The process of implementing inter-domain multicast delivery by leveraging MSDP peers is as follows:

- 1 When the multicast source in PIM-SM 1 sends the first multicast packet to multicast group G, DR 1 encapsulates the multicast data within a register message and sends the register message to RP 1. Then, RP 1 gets aware of the information related to the multicast source.
- 2 As the source-side RP, RP 1 creates SA messages and periodically sends the SA messages to its MSDP peer. An SA message contains the source address (S), the multicast group address (G), and the address of the RP which has created this SA message (namely RP 1).
- 3 On MSDP peers, each SA message is subject to a reverse path forwarding (RPF) check and multicast policy-based filtering, so that only SA messages that have arrived along the correct path and passed the filtering are received and forwarded. This avoids delivery loops of SA messages. In addition, you can configure MSDP peers into an MSDP mesh group so as to avoid flooding of SA messages between MSDP peers.
- 4 SA messages are forwarded from one MSDP peer to another, and finally the information of the multicast source traverses all PIM-SM domains with MSDP peers (PIM-SM 2 and PIM-SM 3 in this example).

- 5 Upon receiving the SA message create by RP 1, RP 2 in PIM-SM 2 checks whether there are any receivers for the multicast group in the domain.
- 6 If so, the RPT for the multicast group G is maintained between RP 2 and the receivers. RP 2 creates an (S, G) entry, and sends an (S, G) join message hop by hop towards DR 1 at the multicast source side, so that it can directly join the SPT rooted at the source over other PIM-SM domains. Then, the multicast data can flow along the SPT to RP 2 and is forwarded by RP 2 to the receivers along the RPT. Upon receiving the multicast traffic, the DR at the receiver side (DR 2) decides whether to initiate an RPT-to-SPT switchover process.
- 7 If no receivers for the group exist in the domain, RP 2 does not create an (S, G) entry and does not join the SPT rooted at the source.



- An MSDP mesh group refers to a group of MSDP peers that have MSDP peering relationships among one another and share the same group name.
- When using MSDP for inter-domain multicasting, once an RP receives information from a multicast source, it no longer relies on RPs in other PIM-SM domains. The receivers can override the RPs in other domains and directly join the multicast source based SPT.

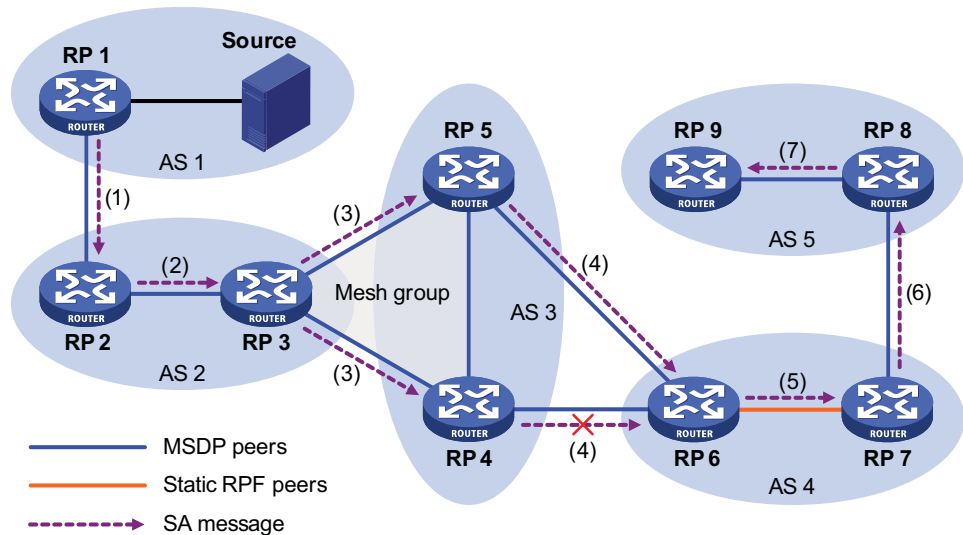
RPF check rules for SA messages

As shown in Figure 201, there are five autonomous systems in the network, AS 1 through AS 5, with IGP enabled on routers within each AS and EBGP as the interoperation protocol among different ASs. Each AS contains at least one PIM-SM domain and each PIM-SM domain contains one or more RPs. MSDP peering relationships have been established among different RPs. RP 3, RP 4 and RP 5 are in an MSDP mesh group. On RP 7, RP 6 is configured as its static RPF peer.



If only one MSDP peer exists in a PIM-SM domain, this PIM-SM domain is also called a stub domain. For example, AS 4 in Figure 201 is a stub domain. The MSDP peer in a stub domain can have multiple remote MSDP peers at the same time. You can configure one or more remote MSDP peers as static RPF peers. When an RP receives an SA message from a static RPF peer, the RP accepts the SA message and forwards it to other peers without performing an RPF check.

Figure 201 Diagram for RPF check for SA messages



As illustrated in Figure 201, these MSDP peers dispose of SA messages according to the following RPF check rules:

- 1 When RP 2 receives an SA message from RP 1
Because the source-side RP address carried in the SA message is the same as the MSDP peer address, which means that the MSDP peer where the SA is from is the RP that has created the SA message, RP 2 accepts the SA message and forwards it to its other MSDP peer (RP 3).
- 2 When RP 3 receives the SA message from RP 2
Because the SA message is from an MSDP peer (RP 2) in the same AS, and the MSDP peer is the next hop on the optimal path to the source-side RP, RP 3 accepts the message and forwards it to other peers (RP 4 and RP 5).
- 3 When RP 4 and RP 5 receive the SA message from RP 3
Because the SA message is from an MSDP peer (RP 3) in the same mesh group, RP 4 and RP 5 both accept the SA message, but they do not forward the message to other members in the mesh group; instead, they forward it to other MSDP peers (RP 6 in this example) out of the mesh group.
- 4 When RP 6 receives the SA messages from RP 4 and RP 5 (suppose RP 5 has a higher IP address)
Although RP 4 and RP 5 are in the same AS (AS 3) and both are MSDP peers of RP 6, because RP 5 has a higher IP address, RP 6 accepts only the SA message from RP 5.
- 5 When RP 7 receives the SA message from RP 6
Because the SA message is from a static RPF peer (RP 6), RP 7 accepts the SA message and forwards it to other peer (RP 8).
- 6 When RP 8 receives the SA message from RP 7
An EBGP route exists between two MSDP peers in different ASs. Because the SA message is from an MSDP peer (RP 7) in a different AS, and the MSDP peer is the next hop on the EBGP route to the source-side RP, RP 8 accepts the message and forwards it to its other peer (RP 9).
- 7 When RP 9 receives the SA message from RP 8
Because RP 9 has only one MSDP peer, RP 9 accepts the SA message.

SA messages from other paths than described above will not be accepted nor forwarded by MSDP peers.

Implementing intra-domain Anycast RP by leveraging MSDP peers

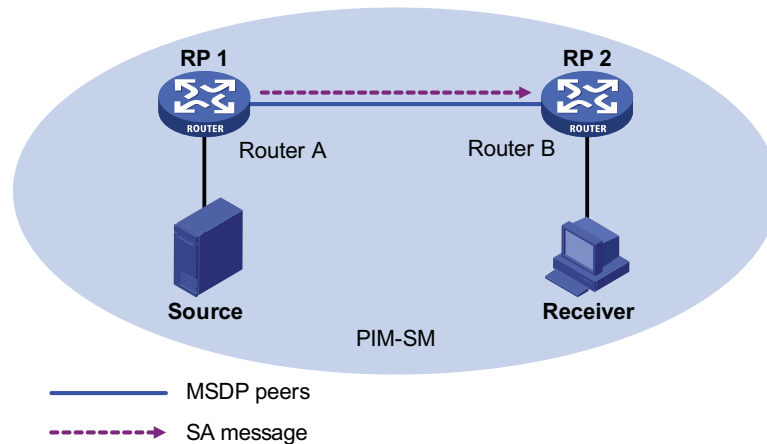
Anycast RP refers to such an application that enables load balancing and redundancy backup between two or more RPs within a PIM-SM domain by configuring the same IP address for, and establishing MSDP peering relationships between, these RPs.

As shown in Figure 202, within a PIM-SM domain, a multicast source sends multicast data to multicast group G, and Receiver is a member of the multicast group. To implement Anycast RP, configure the same IP address (known as anycast RP address, typically a private address) on Router A and Router B, configure these interfaces as C-RPs, and establish an MSDP peering relationship between Router A and Router B.



Usually an Anycast RP address is configured on a logic interface, like a loopback interface.

Figure 202 Typical network diagram of Anycast RP



The work process of Anycast RP is as follows:

- 1 The multicast source registers with the nearest RP. In this example, Source registers with RP 1, with its multicast data encapsulated in the register message. When the register message arrives to RP 1, RP 1 decapsulates the message.
- 2 Receivers send join messages to the nearest RP to join in the RPT rooted as this RP. In this example, Receiver joins the RPT rooted at RP 2.
- 3 RPs share the registered multicast information by means of SA messages. In this example, RP 1 creates an SA message and sends it to RP 2, with the multicast data from Source encapsulated in the SA message. When the SA message reaches RP 2, RP 2 decapsulates the message.
- 4 Receivers receive the multicast data along the RPT and directly join the SPT rooted at the multicast source. In this example, RP 2 forwards the multicast data down the RPT. When Receiver receives the multicast data from Source, it directly joins the SPT rooted at Source.

The significance of Anycast RP is as follows:

- Optimal RP path: A multicast source registers with the nearest RP so that an SPT with the optimal path is built; a receiver joins the nearest RP so that an RPT with the optimal path is built.
- Load balancing between RPs: Each RP just needs to maintain part of the source/group information within the PIM-SM domain and forward part of the multicast data, thus achieving load balancing between different RPs.
- Redundancy backup between RPs: When an RP fails, the multicast source previously registered on it or the receivers previous joined it will register with or join another nearest RP, thus achieving redundancy backup between RPs.



CAUTION:

- Be sure to configure a 32-bit subnet mask (255.255.255.255) for the Anycast RP address, namely configure the Anycast RP address into a host address.
- An MSDP peer address must be different from the Anycast RP address.

- Protocols and Standards** MSDP is documented in the following specifications:
- RFC 3618: Multicast Source Discovery Protocol (MSDP)
 - RFC 3446: Anycast Rendezvous Point (RP) mechanism using Protocol Independent Multicast (PIM) and Multicast Source Discovery Protocol (MSDP)

MSDP Configuration Task List

Complete these tasks to configure MSDP:

Task	Remarks
"Configuring Basic Functions of MSDP" on page 679	"Enabling MSDP" on page 679 Required
	"Creating an MSDP Peer Connection" on page 680 Required
	"Configuring a Static RPF Peer" on page 680 Optional
"Configuring an MSDP Peer Connection" on page 680	"Configuring MSDP Peer Description" on page 681 Optional
	"Configuring an MSDP Mesh Group" on page 681 Optional
	"Configuring MSDP Peer Connection Control" on page 682 Optional
"Configuring SA Messages Related Parameters" on page 682	"Configuring SA Message Content" on page 682 Optional
	"Configuring SA Request Messages" on page 683 Optional
	"Configuring an SA Message Filtering Rule" on page 684 Optional
	"Configuring SA Message Cache" on page 684 Optional

Configuring Basic Functions of MSDP



All the configuration tasks should be carried out on RPs in PIM-SM domains, and each of these RPs acts as an MSDP peer.

Configuration Prerequisites

Before configuring the basic functions of MSDP, complete the following tasks:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Configuring PIM-SM to enable intra-domain multicast forwarding.

Before configuring the basic functions of MSDP, prepare the following data:

- IP addresses of MSDP peers
- Address prefix list for an RP address filtering policy

Enabling MSDP

Follow these steps to enable MSDP:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable IP multicast routing	multicast routing-enable	Required Disabled by default
Enable MSDP and enter MSDP view	msdp	Required Disabled by default

Creating an MSDP Peer Connection

An MSDP peering relationship is identified by an address pair, namely the address of the local MSDP peer and that of the remote MSDP peer. An MSDP peer connection must be created on both devices that are a pair of MSDP peers.

Follow these steps to create an MSDP peer connection:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter MSDP view	msdp	-
Create an MSDP peer connection	peer peer-address connect-interface interface-type interface-number	Required No MSDP peer connection created by default



If an interface of the router is shared by an MSDP peer and a BGP peer at the same time, we recommend that you configuration the same IP address for the MSDP peer and BGP peer.

Configuring a Static RPF Peer

Configuring static RPF peers avoids RPF check of SA messages.

Follow these steps to configure a static RPF peer:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter MSDP view	msdp	-
Configure a static RPF peer	static-rpf-peer peer-address [rp-policy ip-prefix-name]	Required No static RPF peer configured by default



If only one MSDP peer is configured on a router, this MSDP will be registered as a static RPF peer.

Configuring an MSDP Peer Connection

Configuration Prerequisites

Before configuring MSDP peer connection, complete the following tasks:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Configuring basic functions of MSDP

Before configuring an MSDP peer connection, prepare the following data:

- Description information of MSDP peers
- Name of an MSDP mesh group
- MSDP peer connection retry interval

Configuring MSDP Peer Description

With the MSDP peer description information, the administrator can easily distinguish different MSDP peers and thus better manage MSDP peers.

Follow these steps to configure description for an MSDP peer:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter MSDP view	msdp	-
Configure description for an MSDP peer	peer <i>peer-address</i> description <i>text</i>	Required No description for MSDP peers by default

Configuring an MSDP Mesh Group

An AS may contain multiple MSDP peers. You can use the MSDP mesh group mechanism to avoid SA message flooding among these MSDP peers and optimize the multicast traffic.

On one hand, an MSDP peer in an MSDP mesh group forwards SA messages from outside the mesh group that have passed the RPF check to the other members in the mesh group; on the other hand, a mesh group member accepts SA messages from inside the group without performing an RPF check, and does not forward the message within the mesh group either. This mechanism not only avoids SA flooding but also simplifies the RPF check mechanism, because BGP is not needed to run between these MSDP peers.

By configuring the same mesh group name for multiple MSDP peers, you can create a mesh group with these MSDP peers.

Follow these steps to create an MSDP mesh group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter MSDP view	msdp	-
Create an MSDP peer as a mesh group member	peer <i>peer-address</i> mesh-group <i>name</i>	Required An MSDP peer does not belong to any mesh group by default



- *Before grouping multiple routers into an MSDP mesh group, make sure that these routers are interconnected with one another.*
- *If you configure more than one mesh group name on an MSDP peer, only the last configuration is effective.*

Configuring MSDP Peer Connection Control

MSDP peers are interconnected over TCP (port number 639). You can flexibly control sessions between MSDP peers by manually deactivating and reactivating the MSDP peering connections. When the connection between two MSDP peers is deactivated, SA messages will no longer be delivered between them, and the TCP connection is closed without any connection setup retry, but the configuration information will remain unchanged.

When a new MSDP peer is created, or when a previously deactivated MSDP peer connection is reactivated, or when a previously failed MSDP peer attempts to resume operation, a TCP connection is required. You can flexibly adjust the interval between MSDP peering connection retries.

Follow these steps to configure MSDP peer connection control:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter MSDP view	msdp	-
Deactivate an MSDP peer	shutdown <i>peer-address</i>	Optional Active by default
Configure the interval between MSDP peer connection retries	timer retry <i>interval</i>	Optional 30 seconds by default

Configuring SA Messages Related Parameters

Configuration Prerequisites

Before configuring SA message delivery, complete the following tasks:

- Configure any unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Configuring basic functions of MSDP

Before configuring SA message delivery, prepare the following data:

- ACL as a filtering rule for SA request messages
- ACL as an SA message creation rule
- ACL as a filtering rule for receiving or forwarding SA messages
- Minimum TTL value of multicast packets encapsulated in SA messages
- Maximum SA message cache size

Configuring SA Message Content

Some multicast sources send multicast data at an interval longer than the aging time of (S, G) entries. In this case, the source-side DR has to encapsulate multicast data packet by packet in register messages and send them to the source-side RP. The source-side RP transmits the (S, G) information to the remote RP through SA messages. Then the remote RP joins the source-side DR and builds an SPT. Since the (S, G) entries have timed out, remote receivers can never receive the multicast data from the multicast source.

If the source-side RP is enabled to encapsulate register messages in SA messages, when there is a multicast packet to deliver, the source-side RP encapsulates a register message containing the multicast packet in an SA message and sends it out. After receiving the SA message, the remote RP decapsulates the SA message and delivers the multicast data contained in the register message to the receivers along the RPT.

The MSDP peers deliver SA messages to one another. Upon receiving an SA message, a router performs RPF check on the message. If the router finds that the remote RP address is the same as the local RP address, it will discard the SA message. In the Anycast RP application, however, you need to configure RPs with the same IP address on two or more routers in the same PIM-SM domain, and configure these routers as MSDP peers to one another. Therefore, a logic RP address (namely the RP address on the logic interface) that is different from the actual RP address must be designated for SA messages so that the messages can pass the RPF check.

Follow these steps to configure the SA message content:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter MSDP view	msdp	-
Enable encapsulation of a register message	encap-data-enable	Optional Disabled by default
Configure the interface address as the RP address in SA messages	originating-rp <i>interface-type interface-number</i>	Optional PIM RP address by default

Configuring SA Request Messages

By default, upon receiving a new Join message, a router does not send an SA request message to its designated MSDP peer; instead, it waits for the next SA message from its MSDP peer. This will cause the receiver to delay obtaining multicast source information. To enable a new receiver to get the currently active multicast source information as early as possible, you can configure routers to send SA request messages to the designated MSDP peers upon receiving a Join message of a new receiver.

Follow these steps to configure SA message transmission and filtering:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter MSDP view	msdp	-
Enable the device to send SA request messages	peer <i>peer-address</i> request-sa-enable	Optional Disabled by default
Configure a filtering rule for SA request messages	peer <i>peer-address</i> sa-request-policy [acl <i>acl-number</i>]	Optional SA request messages are not filtered by default



CAUTION: Before you can enable the device to send SA requests, be sure to disable the SA message cache mechanism.

Configuring an SA Message Filtering Rule

By configuring an SA message creation rule, you can enable the router to filter the (S, G) entries to be advertised when creating an SA message, so that the propagation of messages of multicast sources is controlled.

In addition to controlling SA message creation, you can also configure filtering rules for forwarding and receiving SA messages, so as to control the propagation of multicast source information in the SA messages.

- By configuring a filtering rule for receiving or forwarding SA messages, you can enable the router to filter the (S, G) forwarding entries to be advertised when receiving or forwarding an SA message, so that the propagation of multicast source information is controlled at SA message reception or forwarding.
- An SA message with encapsulated multicast data can be forwarded to a designated MSDP peer only if the TTL value in its IP header exceeds the threshold. Therefore, you can control the forwarding of such an SA message by configuring the TTL threshold of the encapsulated data packet.

Follow these steps to configure a filtering rule for receiving or forwarding SA messages:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter MSDP view	msdp	-
Configure an SA message creation rule	import-source [acl <i>acl-number</i>]	Required No restrictions on (S, G) entries by default
Configure a filtering rule for receiving or forwarding SA messages	peer <i>peer-address</i> sa-policy { import export } [acl <i>acl-number</i>]	Required No filtering rule by default
Configure the minimum TTL value of multicast packets to be encapsulated in SA messages	peer <i>peer-address</i> minimum-ttl <i>tvl-value</i>	Optional 0 by default

Configuring SA Message Cache

To reduce the time spent in obtaining the multicast source information, you can have SA messages cached on the router. However, the more SA messages are cached, the larger memory space of the router is used.

With the SA cache mechanism enabled, when receiving a new Join message, the router will not send an SA request message to its MSDP peer; instead, it acts as follows:

- If there is no SA message in the cache, the router will wait for the SA message sent by its MSDP peer in the next cycle;
- If there is an SA message in the cache, the router will obtain the information of all active sources directly from the SA message and join the corresponding SPT.

To protect the router against denial of service (DoS) attacks, you can configure the maximum number of SA messages the route can cache.

Follow these steps to configure the SA message cache:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter MSDP view	msdp	-
Enable the SA message cache mechanism	cache-sa-enable	Optional Enabled by default
Configure the maximum number of SA messages the router can cache	peer <i>peer-address</i> sa-cache-maximum <i>sa-limit</i>	Optional 8192 by default

Displaying and Maintaining MSDP

To do...	Use the command...	Remarks
View the brief information of MSDP peers	display msdp brief [state { connect down listen shutdown up }]	Available in any view
View the detailed information about the status of MSDP peers	display msdp peer-status [<i>peer-address</i>]	Available in any view
View the (S, G) entry information in the MSDP cache	display msdp sa-cache [<i>group-address</i> <i>source-address</i> <i>as-number</i>] *	Available in any view
View the number of SA messages in the MSDP cache	display msdp sa-count [<i>as-number</i>]	Available in any view
Reset the TCP connection with an MSDP peer	reset msdp peer [<i>peer-address</i>]	Available in user view
Clear (S, G) entries in the MSDP cache	reset msdp sa-cache [<i>group-address</i>]	Available in user view
Clear all statistics information of an MSDP peer	reset msdp statistics [<i>peer-address</i>]	Available in user view

MSDP Configuration Examples

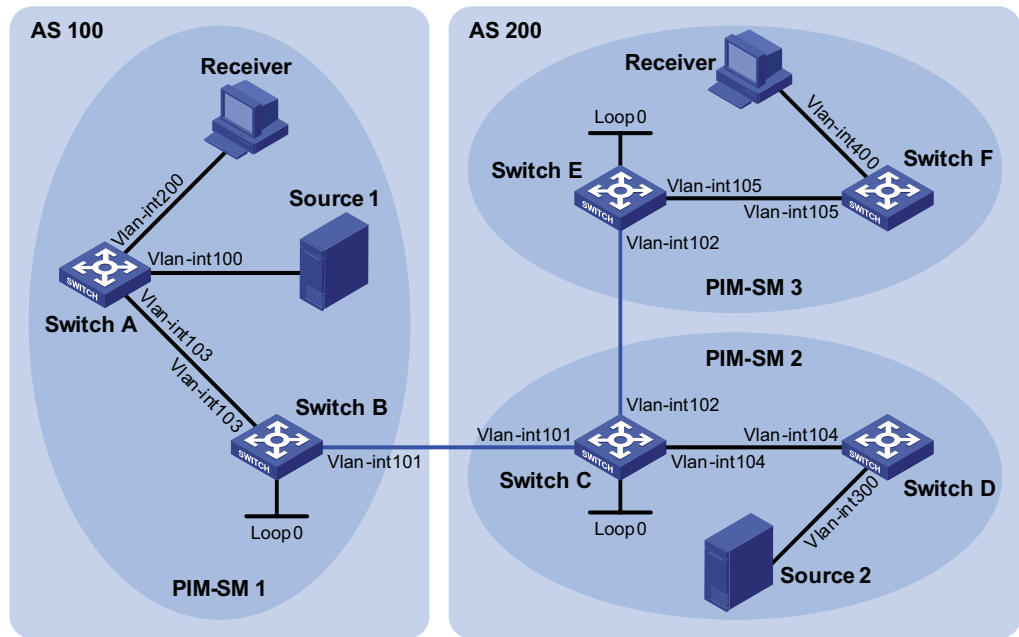
Inter-AS Multicast Configuration Leveraging BGP Routes

Network requirements

- There are two ASs in the network, AS 100 and AS 200 respectively. OSPF is running within each AS, and BGP is running between the two ASs.
- PIM-SM 1 belongs to AS 100, while PIM-SM 2 and PIM-SM 3 belong to AS 200.
- Each PIM-SM domain has zero or one multicast source and receiver. OSPF runs within each domain to provide unicast routes.
- It is required that the respective Loopback 0 of Switch B, Switch C and Switch E be configured as the C-BSR and C-RP of the respective PIM-SM domains.
- It is required that an MSDP peering relationship be set up between Switch B and Switch C through EBGP, and between Switch C and Switch E through IBGP.

Network diagram

Figure 203 Network diagram for inter-AS multicast configuration leveraging BGP routes



MSDP peers

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int103	10.110.1.2/24	Switch D	Vlan-int104	10.110.4.2/24
	Vlan-int100	10.110.2.1/24		Vlan-int300	10.110.5.1/24
	Vlan-int200	10.110.3.1/24		Switch E	Vlan-int105
Switch B	Vlan-int103	10.110.1.1/24	Vlan-int102		192.168.3.2/24
	Vlan-int101	192.168.1.1/24	Loop0	3.3.3.3/32	
	Loop0	1.1.1.1/32	Switch F	Vlan-int105	10.110.6.2/24
Switch C	Vlan-int104	10.110.4.1/24		Vlan-int400	10.110.7.1/24
	Vlan-int102	192.168.3.1/24	Source 1	-	10.110.2.100/24
	Vlan-int101	192.168.1.2/24	Source 2	-	10.110.5.100/24
Loop0	2.2.2.2/32				

Configuration procedure

- 1 Configure the interface IP addresses and unicast routing protocol for each switch. Configure the IP address and subnet mask for each interface as per Figure 203. Detailed configuration steps are omitted here.

Configure OSPF for interconnection between switches in each AS. Ensure the network-layer interoperability among each AS, and ensure the dynamic update of routing information between the switches through a unicast routing protocol. Detailed configuration steps are omitted here.
- 2 Enable IP multicast routing, enable PIM-SM on each interface, and configure a PIM-SM domain border.

Enable IP multicast routing on Switch A, enable PIM-SM on each interface, and enable IGMP on the host-side interface VLAN-interface 200.

```

<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 103
[SwitchA-Vlan-interface103] pim sm
[SwitchA-Vlan-interface103] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] igmp enable
[SwitchA-Vlan-interface200] pim sm
[SwitchA-Vlan-interface200] quit

```

The configuration on Switch B, Switch C, Switch D, Switch E, and Switch F is similar to the configuration on Switch A.

Configure a PIM domain border on Switch B.

```

[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] pim bsr-boundary
[SwitchB-Vlan-interface101] quit

```

The configuration on Switch C and Switch E is similar to the configuration on Switch B.

3 Configure C-BSRs and C-RPs

Configure Loopback 0 as a C-BSR and a C-RP on Switch B.

```

[SwitchB] pim
[SwitchB-pim] c-bsr loopback 0
[SwitchB-pim] c-rp loopback 0
[SwitchB-pim] quit

```

The configuration on Switch C and Switch E is similar to the configuration on Switch B.

4 Configure BGP for mutual route redistribution between BGP and OSPF

Configure EBGP on Switch B, and redistribute OSPF routes.

```

[SwitchB] bgp 100
[SwitchB-bgp] router-id 1.1.1.1
[SwitchB-bgp] peer 192.168.1.2 as-number 200
[SwitchB-bgp] import-route ospf 1
[SwitchB-bgp] quit

```

Configure IBGP and EBGP on Switch C, and redistribute OSPF routes.

```

[SwitchC] bgp 200
[SwitchC-bgp] router-id 2.2.2.2
[SwitchC-bgp] peer 192.168.1.1 as-number 100
[SwitchC-bgp] peer 192.168.3.2 as-number 200
[SwitchC-bgp] import-route ospf 1
[SwitchC-bgp] quit

```

Configure IBGP on Switch E, and redistribute OSPF routes.

```

[SwitchE] bgp 200
[SwitchE-bgp] router-id 3.3.3.3
[SwitchE-bgp] peer 192.168.3.1 as-number 200
[SwitchE-bgp] import-route ospf 1
[SwitchE-bgp] quit

```

Redistribute BGP routes into OSPF on Switch B.

```
[SwitchB] ospf 1
[SwitchB-ospf-1] import-route bgp
[SwitchB-ospf-1] quit
```

The configuration on Switch C and Switch E is similar to the configuration on Switch B.

5 Configure MSDP peers

Configure an MSDP peer on Switch B.

```
[SwitchB] msdp
[SwitchB-msdp] peer 192.168.1.2 connect-interface vlan-interface 101
[SwitchB-msdp] quit
```

Configure an MSDP peer on Switch C.

```
[SwitchC] msdp
[SwitchC-msdp] peer 192.168.1.1 connect-interface vlan-interface 101
[SwitchC-msdp] peer 192.168.3.2 connect-interface vlan-interface 102
[SwitchC-msdp] quit
```

Configure MSDP peers on Switch E.

```
[SwitchE] msdp
[SwitchE-msdp] peer 192.168.3.1 connect-interface vlan-interface 102
[SwitchE-msdp] quit
```

6 Verify the configuration

Carry out the **display bgp peer** command to view the BGP peering relationships between the switches. For example:

View the information about BGP peering relationships on Switch B.

```
[SwitchB] display bgp peer

BGP local router ID : 1.1.1.1
Local AS number : 100
Total number of peers : 1                Peers in established state : 1

Peer          V  AS  MsgRcvd  MsgSent  OutQ  PrefRcv  Up/Down  State
-----
192.168.1.2   4 200      24      21      0        6 00:13:09 Established
```

View the information about BGP peering relationships on Switch C.

```
[SwitchC] display bgp peer

BGP local router ID : 2.2.2.2
Local AS number : 200
Total number of peers : 2                Peers in established state : 2

Peer          V  AS  MsgRcvd  MsgSent  OutQ  PrefRcv  Up/Down  State
-----
192.168.1.1   4 100      18      16      0         1 00:12:04 Established
192.168.3.2   4 200      21      20      0         6 00:12:05 Established
```

View the information about BGP peering relationships on Switch E.

```
[SwitchE] display bgp peer

BGP local router ID : 3.3.3.3
Local AS number : 200
Total number of peers : 1                Peers in established state : 1

Peer          V  AS  MsgRcvd  MsgSent  OutQ  PrefRcv  Up/Down  State
-----
192.168.3.1   4 200      16      14      0         1 00:10:58 Established
```

To view the BGP routing table information on the switches, use the **display bgp routing-table** command. For example:

View the BGP routing table information on Switch C.

```
[SwitchC] display bgp routing-table

Total Number of Routes: 13

BGP Local router ID is 2.2.2.2
Status codes: * - valid, > - best, d - damped,
              h - history, i - internal, s - suppressed, S - Stale
              Origin : i - IGP, e - EGP, ? - incomplete

      Network                NextHop      MED      LocPrf    PrefVal Path/Ogn
*>  1.1.1.1/32              192.168.1.1  0                0      100?
*>i  2.2.2.2/32              192.168.3.2  0                100     0      ?
*>  3.3.3.3/32              0.0.0.0      0                0      0      ?
*>  192.168.1.0             0.0.0.0      0                0      0      ?
*   192.168.1.1             192.168.1.1  0                0      0      100?
*>  192.168.1.1/32          0.0.0.0      0                0      0      ?
*>  192.168.1.2/32          0.0.0.0      0                0      0      ?
*   192.168.1.1             192.168.1.1  0                0      0      100?
*>  192.168.3.0             0.0.0.0      0                0      0      ?
* i  192.168.3.2             192.168.3.2  0                100     0      ?
*>  192.168.3.1/32          0.0.0.0      0                0      0      ?
*>  192.168.3.2/32          0.0.0.0      0                0      0      ?
* i  192.168.3.2             192.168.3.2  0                100     0      ?
```

When the multicast source in PIM-SM 1 (Source 1) and the multicast source in PIM-SM 2 (Source 2) send multicast information, receivers in PIM-SM 1 and PIM-SM 3 can receive the multicast data. You can use the **display msdp brief** command to view the brief information of MSDP peering relationships between the switches. For example:

View the brief information about MSDP peering relationships on Switch B.

```
[SwitchB] display msdp brief
MSDP Peer Brief Information

Configured  Up        Listen    Connect    Shutdown    Down
1           1         0         0          0           0

Peer's Address  State  Up/Down time  AS  SA Count  Reset Count
192.168.1.2    Up     00:12:27     200  13        0
```

View the brief information about MSDP peering relationships on Switch C.

```
[SwitchC] display msdp brief
MSDP Peer Brief Information

Configured  Up        Listen    Connect    Shutdown    Down
2           2         0         0          0           0

Peer's Address  State  Up/Down time  AS  SA Count  Reset Count
192.168.3.2    Up     00:15:32     200  8         0
192.168.1.1    Up     00:06:39     100  13        0
```

View the brief information about MSDP peering relationships on Switch E.

```
[SwitchE] display msdp brief
MSDP Peer Brief Information

Configured  Up        Listen    Connect    Shutdown    Down
1           1         0         0          0           0

Peer's Address  State  Up/Down time  AS  SA Count  Reset Count
192.168.3.1    Up     01:07:08     200  8         0
```

View the detailed MSDP peer information on Switch B.

```
[SwitchB] display msdp peer-status
MSDP Peer 192.168.1.2, AS 200
```

```

Description:
Information about connection status:
  State: Up
  Up/down time: 00:15:47
  Resets: 0
  Connection interface: Vlan-interface101 (192.168.1.1)
  Number of sent/received messages: 16/16
  Number of discarded output messages: 0
  Elapsed time since last connection or counters clear: 00:17:51
Information about (Source, Group)-based SA filtering policy:
  Import policy: none
  Export policy: none
Information about SA-Requests:
  Policy to accept SA-Request messages: none
  Sending SA-Requests status: disable
Minimum TTL to forward SA with encapsulated data: 0
SAs learned from this peer: 0, SA-cache maximum for the peer: none
Input queue size: 0, Output queue size: 0
Counters for MSDP message:
  Count of RPF check failure: 0
  Incoming/outgoing SA messages: 0/0
  Incoming/outgoing SA requests: 0/0
  Incoming/outgoing SA responses: 0/0
  Incoming/outgoing data packets: 0/0

```

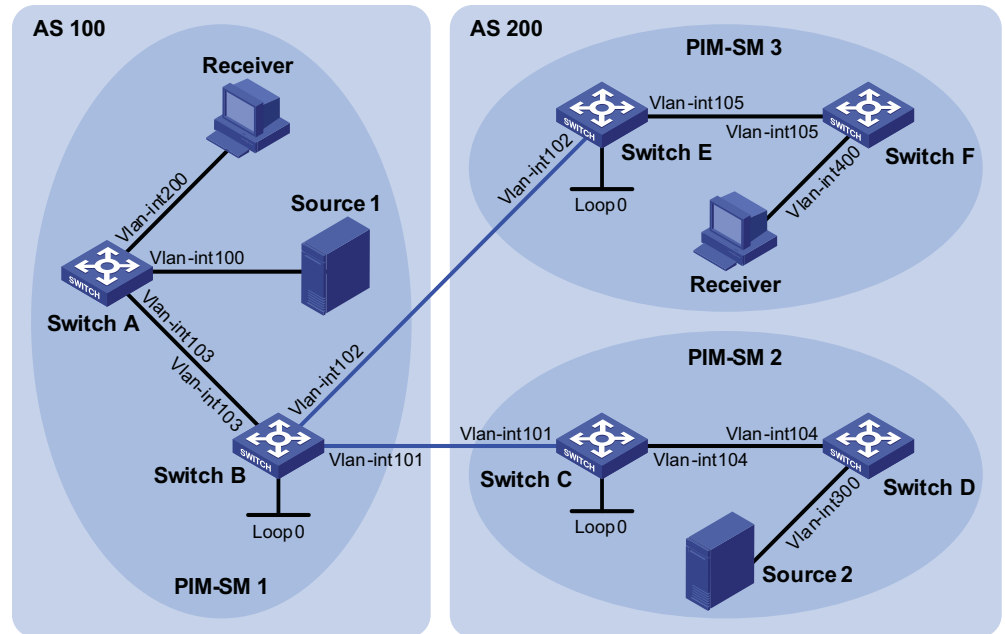
Inter-AS Multicast Configuration Leveraging Static RPF Peers

Network requirements

- There are two ASs in the network, AS 100 and AS 200 respectively. OSPF is running within each AS, and BGP is running between the two ASs.
- PIM-SM 1 belongs to AS 100, while PIM-SM 2 and PIM-SM 3 belong to AS 200.
- Each PIM-SM domain has zero or one multicast source and receiver. OSPF runs within each domain to provide unicast routes.
- PIM-SM 2 and PIM-SM 3 are both stub domains, and BGP or MBGP is not required between these two domains and PIM-SM 1. Instead, static RPF peers are configured to avoid RPF check on SA messages.
- It is required that the respective loopback 0 of Switch B, Switch C and Switch E be configured as the C-BSR and C-RP of the respective PIM-SM domains.
- It is required that Switch C and Switch E be configured as static RPF peers of Switch B, and Switch B be configured as the only static RPF peer of Switch C and Switch E, so that any switch can receive SA messages only from its static RPF peer(s) and permitted by the corresponding filtering policy.

Network diagram

Figure 204 Network diagram for inter-AS multicast configuration leveraging static RPF peers



— Static RPF peers

Device	Interface	IP address	Device	Interface	IP address
Switch A	Vlan-int103	10.110.1.2/24	Switch D	Vlan-int104	10.110.4.2/24
	Vlan-int100	10.110.2.1/24		Vlan-int300	10.110.5.1/24
	Vlan-int200	10.110.3.1/24		Switch E	Vlan-int105
Switch B	Vlan-int103	10.110.1.1/24	Vlan-int102		192.168.3.2/24
	Vlan-int101	192.168.1.1/24	Loop0		3.3.3.3/32
	Vlan-int102	192.168.3.1/24	Switch F	Vlan-int105	10.110.6.2/24
	Loop0	1.1.1.1/32		Vlan-int400	10.110.7.1/24
Switch C	Vlan-int101	192.168.1.2/24	Source 1	-	10.110.2.100/24
	Vlan-int104	10.110.4.1/24	Source 2	-	10.110.5.100/24
	Loop0	2.2.2.2/32			

Configuration procedure

- 1 Configure the interface IP addresses and unicast routing protocol for each switch
Configure the IP address and subnet mask for each interface as per Figure 204. Detailed configuration steps are omitted here.

Configure OSPF for interconnection between the switches. Ensure the network-layer interoperability in each AS, and ensure the dynamic update of routing information among the switches through a unicast routing protocol. Detailed configuration steps are omitted here.

- 2 Enable IP multicast routing, enable PIM-SM and IGMP, and configure a PIM-SM domain border

Enable IP multicast routing on Switch A, enable PIM-SM on each interface, and enable IGMP on the host-side interface VLAN-interface 200.

```

<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 103
[SwitchA-Vlan-interface103] pim sm
[SwitchA-Vlan-interface103] quit
[SwitchA] interface vlan-interface 100
[SwitchA-Vlan-interface100] pim sm
[SwitchA-Vlan-interface100] quit
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] igmp enable
[SwitchA-Vlan-interface200] pim sm
[SwitchA-Vlan-interface200] quit

```

The configuration on Switch B, Switch C, Switch D, Switch E, and Switch F is similar to the configuration on Switch A.

Configure PIM domain borders on Switch B.

```

[SwitchB] interface vlan-interface 102
[SwitchB-Vlan-interface102] pim bsr-boundary
[SwitchB-Vlan-interface102] quit
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] pim bsr-boundary
[SwitchB-Vlan-interface101] quit

```

The configuration on Switch C and Switch E is similar to the configuration on Switch B.

3 Configure C-BSRs and C-RPs

Configure Loopback 0 as a C-BSR and a C-RP on Switch B.

```

[SwitchB] pim
[SwitchB-pim] c-bsr loopback 0
[SwitchB-pim] c-rp loopback 0
[SwitchB-pim] quit

```

The configuration on Switch C and Switch E is similar to the configuration on Switch B.

4 Configure a static RPF peer

Configure Switch C and Switch E as a static RPF peers of Switch B.

```

[SwitchB] ip ip-prefix list-df permit 192.168.0.0 16 greater-equal 1
6 less-equal 32
[SwitchB] msdp
[SwitchB-msdp] peer 192.168.3.1 connect-interface vlan-interface 102
[SwitchB-msdp] peer 192.168.1.2 connect-interface vlan-interface 101
[SwitchB-msdp] static-rpf-peer 192.168.3.1 rp-policy list-df
[SwitchB-msdp] static-rpf-peer 192.168.1.2 rp-policy list-df
[SwitchB-msdp] quit

```

Configure Switch B as a static RPF peer of Switch C.

```

[SwitchC] ip ip-prefix list-c permit 192.168.0.0 16 greater-equal 16
less-equal 32
[SwitchC] msdp
[SwitchC-msdp] peer 192.168.3.2 connect-interface vlan-interface 102
[SwitchC-msdp] static-rpf-peer 192.168.3.2 rp-policy list-c
[SwitchC-msdp] quit

```

Configure Switch B as a static RPF peer of Switch E.


```
[SwitchE] ip ip-prefix list-c permit 192.168.0.0 16 greater-equal 16 less-equal 32
[SwitchE] msdp
[SwitchE-msdp] peer 192.168.3.2 connect-interface vlan-interface 102
[SwitchE-msdp] static-rpf-peer 192.168.3.2 rp-policy list-c
[SwitchE-msdp] quit
```

5 Verify the configuration

Carry out the **display bgp peer** command to view the BGP peering relationships between the switches. If the command gives no output information, a BGP peering relationship has not been established between the switches.

When the multicast source in PIM-SM 1 (Source 1) and the multicast source in PIM-SM 2 (Source 2) send multicast information, receivers in PIM-SM 1 and PIM-SM 3 can receive the multicast data. You can use the **display msdp brief** command to view the brief information of MSDP peering relationships between the switches. For example:

View the brief MSDP peer information on Switch B.

```
[SwitchB] display msdp brief
MSDP Peer Brief Information
  Configured   Up           Listen       Connect      Shutdown     Down
  2            2           0           0           0           0

  Peer's Address  State  Up/Down time  AS  SA Count  Reset Count
  192.168.3.2    Up     01:07:08     ?   8         0
  192.168.1.2    Up     00:16:39     ?   13        0
```

View the brief MSDP peer information on Switch C.

```
[SwitchC] display msdp brief
MSDP Peer Brief Information
  Configured   Up           Listen       Connect      Shutdown     Down
  1            1           0           0           0           0

  Peer's Address  State  Up/Down time  AS  SA Count  Reset Count
  192.168.1.1    Up     01:07:09     ?   8         0
```

View the brief MSDP peer information on Switch E.

```
[SwitchE] display msdp brief
MSDP Peer Brief Information
  Configured   Up           Listen       Connect      Shutdown     Down
  1            1           0           0           0           0

  Peer's Address  State  Up/Down time  AS  SA Count  Reset Count
  192.168.3.1    Up     00:16:40     ?   13        0
```

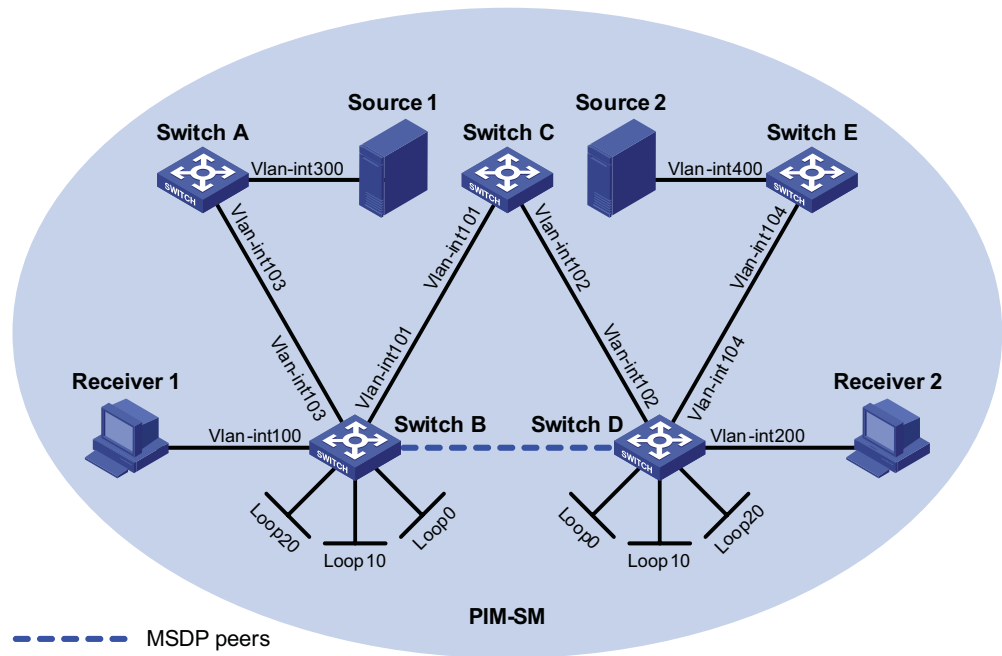
Anycast RP Configuration

Network requirements

- The PIM-SM domain has multiple multicast sources and receivers. OSPF runs within the domain to provide unicast routes.
- It is required to configure the anycast RP application so that the receiver-side DRs and the source-side DRs can initiate a Join message to their respective RPs that are the topologically nearest to them.
- On Switch B and Switch D, configure the interface Loopback 10 as a C-BSR, and Loopback 20 as a C-RP.
- The router ID of Switch B is 1.1.1.1, while the router ID of Switch D is 2.2.2.2. Set up an MSDP peering relationship between Switch B and Switch D.

Network diagram

Figure 205 Network diagram for anycast RP configuration



Device	Interface	IP address	Device	Interface	IP address
Source 1	-	10.110.5.100/24	Switch C	Vlan-int101	192.168.1.2/24
Source 2	-	10.110.6.100/24		Vlan-int102	192.168.2.2/24
Switch A	Vlan-int300	10.110.5.1/24	Switch D	Vlan-int200	10.110.3.1/24
	Vlan-int103	10.110.2.2/24		Vlan-int104	10.110.4.1/24
Switch B	Vlan-int100	10.110.1.1/24		Vlan-int102	192.168.2.1/24
	Vlan-int103	10.110.2.1/24		Loop0	2.2.2.2/32
	Vlan-int101	192.168.1.1/24		Loop10	4.4.4.4/32
	Loop0	1.1.1.1/32		Loop20	10.1.1.1/32
	Loop10	3.3.3.3/32	Switch E	Vlan-int400	10.110.6.1/24
	Loop20	10.1.1.1/32		Vlan-int104	10.110.4.2/24

Configuration procedure

- 1 Configure the interface IP addresses and unicast routing protocol for each switch. Configure the IP address and subnet mask for each interface as per Figure 205. Detailed configuration steps are omitted here.

Configure OSPF for interconnection between the switches. Ensure the network-layer interoperability among the switches, and ensure the dynamic update of routing information between the switches through a unicast routing protocol. Detailed configuration steps are omitted here.

- 2 Enable IP multicast routing, and enable PIM-SM and IGMP

Enable IP multicast routing on Switch B, enable PIM-SM on each interface, and enable IGMP on the host-side interface VLAN-interface100.

```
<SwitchB> system-view
[SwitchB] multicast routing-enable
```

```
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] igmp enable
[SwitchB-Vlan-interface100] pim sm
[SwitchB-Vlan-interface100] quit
[SwitchB] interface vlan-interface 103
[SwitchB-Vlan-interface103] pim sm
[SwitchB-Vlan-interface103] quit
[SwitchB] interface Vlan-interface 101
[SwitchB-Vlan-interface101] pim sm
[SwitchB-Vlan-interface101] quit
[SwitchB] interface loopback 0
[SwitchB-LoopBack0] pim sm
[SwitchB-LoopBack0] quit
[SwitchB] interface loopback 10
[SwitchB-LoopBack10] pim sm
[SwitchB-LoopBack10] quit
[SwitchB] interface loopback 20
[SwitchB-LoopBack20] pim sm
[SwitchB-LoopBack20] quit
```

The configuration on Switch A, Switch C, Switch D, and Switch E is similar to the configuration on Switch B.

3 Configure C-BSRs and C-RPs

Configure Loopback 10 as a C-BSR and Loopback 20 as a C-RP on Switch B.

```
[SwitchB] pim
[SwitchB-pim] c-bsr loopback 10
[SwitchB-pim] c-rp loopback 20
[SwitchB-pim] quit
```

The configuration on Switch D is similar to the configuration on Switch B.

4 Configure MSDP peers

Configure an MSDP peer on Loopback 0 of Switch B.

```
[SwitchB] msdp
[SwitchB-msdp] originating-rp loopback 0
[SwitchB-msdp] peer 2.2.2.2 connect-interface loopback 0
[SwitchB-msdp] quit
```

Configure an MSDP peer on Loopback 0 of Switch D.

```
[SwitchD] msdp
[SwitchD-msdp] originating-rp loopback 0
[SwitchD-msdp] peer 1.1.1.1 connect-interface loopback 0
[SwitchD-msdp] quit
```

5 Verify the configuration

You can use the **display msdp brief** command to view the brief information of MSDP peering relationships between the switches.

View the brief MSDP peer information on Switch B.

```
[SwitchB] display msdp brief
MSDP Peer Brief Information
```

Configured	Up	Listen	Connect	Shutdown	Down
1	1	0	0	0	0

Peer's Address	State	Up/Down time	AS	SA Count	Reset Count
2.2.2.2	Up	00:10:17	?	0	0

View the brief MSDP peer information on Switch D.

```
[SwitchD] display msdp brief
MSDP Peer Brief Information
Configured  Up      Listen      Connect     Shutdown    Down
1           1         0           0           0           0

Peer's Address  State  Up/Down time  AS  SA Count  Reset Count
1.1.1.1        Up     00:10:18     ?   0         0
```

To view the PIM routing information on the switches, use the **display pim routing-table** command. When Source 1 (10.110.5.100/24) sends multicast data to multicast group G (225.1.1.1), Receiver 1 joins multicast group G. By comparing the PIM routing information displayed on Switch B with that displayed on Switch D, you can see that Switch B acts now as the RP for Source 1 and Receiver 1.

View the PIM routing information on Switch B.

```
[SwitchB] display pim routing-table
Total 1 (*, G) entry; 1 (S, G) entry

(*, 225.1.1.1)
  RP: 10.1.1.1 (local)
  Protocol: pim-sm, Flag: WC
  UpTime: 00:15:04
  Upstream interface: Register
    Upstream neighbor: NULL
    RPF prime neighbor: NULL
  Downstream interface(s) information:
  Total number of downstreams: 1
    1: Vlan-interface100
      Protocol: igmp, UpTime: 00:15:04, Expires: -

(10.110.5.100, 225.1.1.1)
  RP: 10.1.1.1 (local)
  Protocol: pim-sm, Flag: SPT 2MSDP ACT
  UpTime: 00:46:28
  Upstream interface: Vlan-interface103
    Upstream neighbor: 10.110.2.2
    RPF prime neighbor: 10.110.2.2
  Downstream interface(s) information:
  Total number of downstreams: 1
    1: Vlan-interface100
      Protocol: pim-sm, UpTime: - , Expires: -
```

View the PIM routing information on Switch D.

```
[SwitchD] display pim routing-table
```

No information is output on Switch D.

Receiver 1 has left multicast group G. Source 1 has stopped sending multicast data to multicast group G. When Source 2 (10.110.6.100/24) sends multicast data to G, Receiver 2 joins G. By comparing the PIM routing information displayed on Switch B with that displayed on Switch D, you can see that Switch D acts now as the RP for Source 2 and Receiver 2.

View the PIM routing information on Switch B.

```
[SwitchB] display pim routing-table
```

No information is output on Switch B.

View the PIM routing information on Switch D.

```
[SwitchD] display pim routing-table
Total 1 (*, G) entry; 1 (S, G) entry

(*, 225.1.1.1)
  RP: 10.1.1.1 (local)
  Protocol: pim-sm, Flag: WC
  UpTime: 00:12:07
  Upstream interface: Register
    Upstream neighbor: NULL
    RPF prime neighbor: NULL
  Downstream interface(s) information:
  Total number of downstreams: 1
    1: Vlan-interface200
      Protocol: igmp, UpTime: 00:12:07, Expires: -

(10.110.6.100, 225.1.1.1)
  RP: 10.1.1.1 (local)
  Protocol: pim-sm, Flag: SPT 2MSDP ACT
  UpTime: 00:40:22
  Upstream interface: Vlan-interface104
    Upstream neighbor: 10.110.4.2
    RPF prime neighbor: 10.110.4.2
  Downstream interface(s) information:
  Total number of downstreams: 1
    1: Vlan-interface200
      Protocol: pim-sm, UpTime: - , Expires: -
```

Troubleshooting MSDP

MSDP Peers Stay in Down State

Symptom

The configured MSDP peers stay in the down state.

Analysis

- A TCP connection-based MSDP peering relationship is established between the local interface address and the MSDP peer after the configuration.
- The TCP connection setup will fail if there is a consistency between the local interface address and the MSDP peer address configured on the router.
- If no route is available between the MSDP peers, the TCP connection setup will also fail.

Solution

- 1 Check that a route is available between the routers. Carry out the **display ip routing-table** command to check whether the unicast route between the routers is correct.
- 2 Check that a unicast route is available between the two routers that will become MSDP peers to each other.
- 3 Verify the interface address consistency between the MSDP peers. Use the **display current-configuration** command to verify that the local interface address and the MSDP peer address of the remote router are the same.

No SA Entries in the Router's SA Cache**Symptom**

MSDP fails to send (S, G) entries through SA messages.

Analysis

- The **import-source** command is used to control sending (S, G) entries through SA messages to MSDP peers. If this command is executed without the *acl-number* argument, all the (S, G) entries will be filtered off, namely no (S, G) entries of the local domain will be advertised.
- If the **import-source** command is not executed, the system will advertise all the (S, G) entries of the local domain. If MSDP fails to send (S, G) entries through SA messages, check whether the **import-source** command has been correctly configured.

Solution

- 1 Check that a route is available between the routers. Carry out the **display ip routing-table** command to check whether the unicast route between the routers is correct.
- 2 Check that a unicast route is available between the two routers that will become MSDP peers to each other.
- 3 Check configuration of the **import-source** command and its *acl-number* argument and make sure that ACL rule can filter appropriate (S, G) entries.

Inter-RP Communication Faults in Anycast RP Application**Symptom**

RPs fail to exchange their locally registered (S, G) entries with one another in the Anycast RP application.

Analysis

- In the Anycast RP application, RPs in the same PIM-SM domain are configured to be MSDP peers to achieve load balancing among the RPs.
- An MSDP peer address must be different from the anycast RP address, and the C-BSR and C-RP must be configured on different devices or interfaces.
- If the **originating-rp** command is executed, MSDP will replace the RP address in the SA messages with the address of the interface specified in the command.
- When an MSDP peer receives an SA message, it performs RPF check on the message. If the MSDP peer finds that the remote RP address is the same as the local RP address, it will discard the SA message.

Solution

- 1 Check that a route is available between the routers. Carry out the **display ip routing-table** command to check whether the unicast route between the routers is correct.
- 2 Check that a unicast route is available between the two routers that will become MSDP peer to each other.
- 3 Check the configuration of the **originating-rp** command. In the Anycast RP application environment, be sure to use the **originating-rp** command to configure the RP address in the SA messages, which must be the local interface address.

- 4 Verify that the C-BSR address is different from the anycast RP address.

49

MULTICAST ROUTING AND FORWARDING CONFIGURATION

When configuring multicast routing and forwarding, go to these sections for information you are interested in:

- “Multicast Routing and Forwarding Overview” on page 701
- “Configuring Multicast Routing and Forwarding” on page 706
- “Displaying and Maintaining Multicast Routing and Forwarding” on page 709
- “Configuration Examples” on page 709
- “Troubleshooting Multicast Routing and Forwarding” on page 713



The term “router” in this document refers to a router in a generic sense or a Layer 3 switch running an IP routing protocol.

Multicast Routing and Forwarding Overview

Introduction to Multicast Routing and Forwarding

In multicast implementations, multicast routing and forwarding are implemented by three types of tables:

- Each multicast routing protocol has its own multicast routing table, such as PIM routing table.
- The information of different multicast routing protocols forms a general multicast routing table.
- The multicast forwarding table is directly used to control the forwarding of multicast packets.

A multicast forwarding table consists of a set of (S, G) entries, each indicating the routing information for delivering multicast data from a multicast source to a multicast group. If a router supports multiple multicast protocols, its multicast routing table will include routes generated by multiple protocols. The router chooses the optimal route from the multicast routing table based on the configured multicast routing and forwarding policy and installs the route entry into its multicast forwarding table.

RPF Mechanism

When creating multicast routing table entries, a multicast routing protocol uses the reverse path forwarding (RPF) mechanism to ensure multicast data delivery along the correct path.

The RPF mechanism enables routers to correctly forward multicast packets based on the multicast route configuration. In addition, the RPF mechanism also helps avoid data loops caused by various reasons.

Implementation of the RPF mechanism

Upon receiving a multicast packet that a multicast source *S* sends to a multicast group *G*, the router first searches its multicast forwarding table:

- 1 If the corresponding (*S*, *G*) entry exists, and the interface on which the packet actually arrived is the incoming interface in the multicast forwarding table, the router forwards the packet to all the outgoing interfaces.
- 2 If the corresponding (*S*, *G*) entry exists, but the interface on which the packet actually arrived is not the incoming interface in the multicast forwarding table, the multicast packet is subject to an RPF check.
- 3 If the result of the RPF check shows that the RPF interface is the incoming interface of the existing (*S*, *G*) entry, this means that the (*S*, *G*) entry is correct but the packet arrived from a wrong path. The packet is to be discarded.
- 4 If the result of the RPF check shows that the RPF interface is not the incoming interface of the existing (*S*, *G*) entry, this means that the (*S*, *G*) entry is no longer valid. The router replaces the incoming interface of the (*S*, *G*) entry with the interface on which the packet actually arrived and forwards the packet to all the outgoing interfaces.
- 5 If no corresponding (*S*, *G*) entry exists in the multicast forwarding table, the packet is also subject to an RPF check. The router creates an (*S*, *G*) entry based on the relevant routing information and using the RPF interface as the incoming interface, and installs the entry into the multicast forwarding table.
- 6 If the interface on which the packet actually arrived is the RPF interface, the RPF check is successful and the router forwards the packet to all the outgoing interfaces.
- 7 If the interface on which the packet actually arrived is not the RPF interface, the RPF check fails and the router discards the packet.

RPF check

The basis for an RPF check is a unicast route or a multicast static route. A unicast routing table contains the shortest path to each destination subnet, while a multicast static routing table lists the RPF routing information defined by the user through static configuration. A multicast routing protocol does not independently maintain any type of unicast route; instead, it relies on the existing unicast routing information or multicast static routes in creating multicast routing entries.

When performing an RPF check, a router searches its unicast routing table and multicast static routing table at the same time. The specific process is as follows:

- 1 The router first chooses an optimal route from the unicast routing table and multicast static routing table:
- 2 The router automatically chooses an optimal unicast route by searching its unicast routing table, using the IP address of the "packet source" as the destination address. The outgoing interface in the corresponding routing entry is the RPF interface and the next hop is the RPF neighbor. The router considers the path along which the packet from the RPF neighbor arrived on the RPF interface to be the shortest path that leads back to the source.
- 3 The router automatically chooses an optimal multicast static route by searching its multicast static routing table, using the IP address of the "packet source" as the

destination address. The corresponding routing entry explicitly defines the RPF interface and the RPF neighbor.

- 4 Then, the router selects one from these two optimal routes as the RPF route. The selection is as follows:
- 5 If configured to use the longest match principle, the router selects the longest match route from the two; if these two routes have the same mask, the route selects the route with a higher priority; if the two routes have the same priority, the router selects the multicast static route.
- 6 If not configured to use the longest match principle, the router selects the route with a higher priority; if the two routes have the same priority, the router selects the multicast static route.



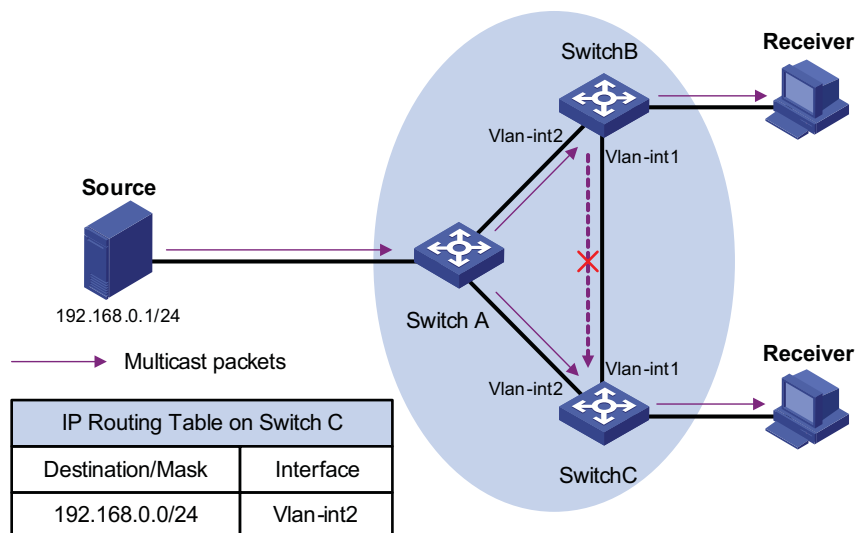
The above-mentioned “packet source” can mean different things in different situations

- For a packet traveling along the shortest path tree (SPT) from the multicast source to the receivers or the source-based tree from the multicast source to the rendezvous point (RP), “packet source” means the multicast source.
- For a packet traveling along the rendezvous point tree (RPT) from the RP to the receivers, “packet source” means the RP.
- For a bootstrap message from the bootstrap router (BSR), “packet source” means the BSR.

For details about the concepts of SPT, RPT and BSR, refer to “PIM Configuration” on page 629.

Assume that unicast routes exist in the network and no multicast static routes have been configured on Switch C, as shown in Figure 206. Multicast packets travel along the SPT from the multicast source to the receivers.

Figure 206 RPF check process



- A multicast packet from Source arrives on VLAN-interface 1 of Switch C, and the corresponding forwarding entry does not exist in the multicast forwarding table of Switch C. Switch C performs an RPF check, and finds in its unicast routing table that the outgoing interface to 192.168.0.0/24 is VLAN-interface

2. This means that the interface on which the packet actually arrived is not the RPF interface. The RPF check fails and the packet is discarded.

- A multicast packet from Source arrives on VLAN-interface 2 of Switch C, and the corresponding forwarding entry does not exist in the multicast forwarding table of Switch C. The switch performs an RPF check, and finds in its unicast routing table that the outgoing interface to 192.168.0.0/24 is the interface on which the packet actually arrived. The RPF check succeeds and the packet is forwarded.

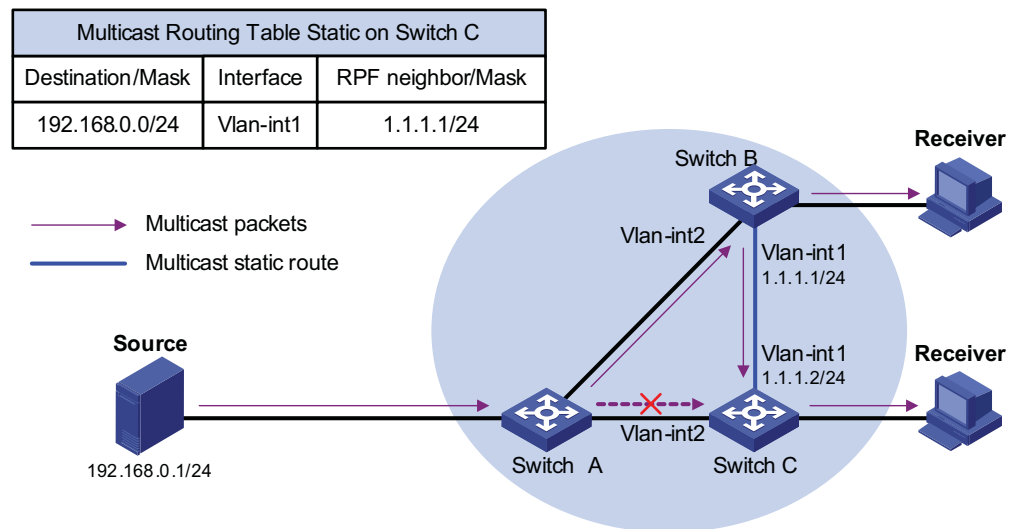
Multicast Static Routes

If the topology structure of a multicast network is the same as that of a unicast network, receivers can receive multicast data via unicast routes. However, the topology structure of a multicast network may differ from that of a unicast network, and some routers may support only unicast but not multicast. In this case, you can configure multicast static routes to provide multicast transmission paths that are different from those for unicast traffic. Note the following two points:

- A multicast static route only affects RPF checks, and not guides multicast forwarding, so it is also called an RPF static route.
- A multicast static route is effective on the multicast router on which it is configured, and will not be broadcast throughout the network or injected to other routers.

A multicast static route is an important basis for RPF checks. With a multicast static route configured on a router, the router searches the unicast routing table and the multicast static routing table simultaneously in a RPF check, chooses the optimal unicast RPF route and the optimal multicast static route respectively from the routing tables, and uses one of them as the RPF route after comparison.

Figure 207 Multicast static route



As shown in Figure 207, when no multicast static route is configured, Switch C's RPF neighbor on the path back to Source is Switch A and the multicast information from Source travels along the path from Switch A to Switch C, which is the unicast route between the two switches; with a static route configured on Switch C and Switch B as Switch C's RPF neighbor on the path back to Source, the

multicast information from Source travels from Switch A to Switch B and then to Switch C.

Multicast Traceroute The multicast traceroute utility is used to trace the path that a multicast stream flows down from the multicast source to the last-hop router.

Concepts in multicast traceroute

- 1 Last-hop router: If a router has one of its interfaces connecting to the subnet the given destination address is on, and if the router is able to forward multicast streams from the given multicast source onto that subnet, that router is called last-hop router.
- 2 First-hop router: the router that directly connects to the multicast source.
- 3 Querier: the router requesting the multicast traceroute.

Introduction to multicast traceroute packets

A multicast traceroute packet is a special IGMP packet, which differs from common IGMP packets in that its IGMP Type field is set to 0x1F or 0x1E and that its destination IP address is a unicast address. There are three types of multicast traceroute packets:

- Query, with the IGMP Type field set to 0x1F,
- Request, with the IGMP Type field set to 0x1F, and
- Response, with the IGMP Type field set to 0x1E.

Process of multicast traceroute

- 1 The querier sends a query to the last-hop router.
- 2 Upon receiving the query, the last-hop router turns the query packet into a request packet by adding a response data block containing its interface addresses and packet statistics to the end of the packet, and forwards the request packet via unicast to the previous hop for the given multicast source and group.
- 3 From the last-hop router to the multicast source, each hop adds a response data block to the end of the request packet and unicasts it to the previous hop.
- 4 When the first-hop router receives the request packet, it changes the packet type to indicate a response packet, and then sends the completed packet via unicast to the multicast traceroute querier.

Configuration Task List

Complete these tasks to configure multicast routing and forwarding:

Task	Remarks
"Enabling IP Multicast Routing" on page 706	Required
"Configuring Multicast Static Routes" on page 706	Optional
"Configuring a Multicast Route Match Rule" on page 707	Optional
"Configuring Multicast Load Splitting" on page 707	Optional
"Configuring a Multicast Forwarding Range" on page 707	Optional
"Configuring the Multicast Forwarding Table Size" on page 708	Optional
"Tracing a Multicast Path" on page 708	Optional

Configuring Multicast Routing and Forwarding

Configuration Prerequisites

Before configuring multicast routing and forwarding, complete the following tasks:

- Configure a unicast routing protocol so that all devices in the domain are interoperable at the network layer.
- Enable PIM (PIM-DM or PIM-SM).

Before configuring multicast routing and forwarding, prepare the following data:

- The minimum TTL value required for a multicast packet to be forwarded
- The maximum number of downstream nodes for a single route in a multicast forwarding table
- The maximum number of routing entries in a multicast forwarding table

Enabling IP Multicast Routing

Before configuring any Layer 3 multicast functionality, you must enable IP multicast routing.

Follow these steps to enable IP multicast routing:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable IP multicast routing	multicast routing-enable	Required Disable by default



CAUTION: IP multicast does not support the use of secondary IP address segments. Namely, multicast can be routed and forwarded only through primary IP addresses, rather than secondary addresses, even if configured on interfaces.

For details about primary and secondary IP addresses, refer to “IP Addressing Configuration” on page 121.

Configuring Multicast Static Routes

Based on the application environment, a multicast static route has the following two functions:

- Changing an RPF route. If the multicast topology structure is the same as the unicast topology in a network, the delivery path of multicast traffic is the same as in unicast. By configuring a multicast static route, you can change the RPF route so as to create a transmission path that is different from the unicast traffic transmission path.
- Creating an RPF route. When a unicast route is interrupted, multicast traffic forwarding is stopped due to lack of an RPF route. By configuring a multicast static route, you can create an RPF route so that a multicast routing entry is created to guide multicast traffic forwarding.

Follow these steps to configure a multicast static route:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure a multicast static route	ip rpf-route-static <i>source-address</i> { <i>mask</i> <i>mask-length</i> } [<i>protocol</i> [<i>process-id</i>]] [route-policy <i>policy-name</i>] { <i>rpf-nbr-address</i> <i>interface-type interface-number</i> } [preference <i>preference</i>] [order <i>order-number</i>]	Required No multicast static route configured by default.



CAUTION: When configuring a multicast static route, you cannot designate an RPF neighbor by specifying an interface (by means of the *interface-type interface-number* command argument combination) if the interface type of that switch is Loopback or VLAN-interface; instead, you can designate an RPF neighbor only by specifying an address (*rpf-nbr-address*).

Configuring a Multicast Route Match Rule

If more than one route exists to the same subnet, a router chooses a route based on the sequence of route configuration.

Follow these steps to configure a multicast route match rule:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the device to select a route based on the longest match	multicast longest-match	Required In order of routing table entries by default

Configuring Multicast Load Splitting

With the load splitting feature enabled, multicast traffic will be evenly distributed among different routes.

Follow these steps to configure multicast load splitting:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configuring multicast load splitting	multicast load-splitting { source source-group }	Required Disabled by default

Configuring a Multicast Forwarding Range

Multicast packets do not travel without a boundary in a network. The multicast data corresponding to each multicast group must be transmitted within a definite scope.

You can configure a forwarding boundary specific to a particular multicast group on all interfaces that support multicast forwarding. A multicast forwarding boundary sets the boundary condition for the multicast groups in the specified range. If the destination address of a multicast packet matches the set boundary condition, the packet will not be forwarded. Once a multicast boundary is configured on an interface, this interface can no longer forward multicast packets (including packets sent from the local device) or receive multicast packets.

Follow these steps to configure a multicast forwarding range:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure a multicast forwarding boundary	multicast boundary <i>group-address</i> { <i>mask</i> <i>mask-length</i> }	Required No forwarding boundary by default

Configuring the Multicast Forwarding Table Size

Too many multicast routing entries can exhaust the router's memory and thus result in lower router performance. Therefore, the number of multicast routing entries should be limited. You can set a limit on the number of entries in the multicast routing table based on the actual networking situation and the performance requirements. In any case, the number of route entries must not exceed the maximum number allowed by the system. This maximum value varies with different device models.

If the configured maximum number of downstream nodes (namely, the maximum number of outgoing interfaces) for a routing entry in the multicast forwarding table is smaller than the current number, the downstream nodes in excess of the configured limit will not be deleted immediately; instead they must be deleted by the multicast routing protocol. In addition, newly added downstream nodes cannot be installed to the routing entry into the forwarding table.

If the configured maximum number of routing entries in the multicast forwarding table is smaller than the current number, the routes in excess of the configured limit will not be deleted immediately; instead they must be deleted by the multicast routing protocol. In addition, newly added route entries cannot be installed to the forwarding table.

Follow these steps to configure the multicast forwarding table size:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the maximum number of downstream nodes for a single route in the multicast forwarding table	multicast forwarding-table downstream-limit <i>limit</i>	Optional The default is 128.
Configure the maximum number of routing entries in the multicast forwarding table	multicast forwarding-table route-limit <i>limit</i>	Optional The default is 1024.

Tracing a Multicast Path

You can run the **mtracert** command to trace the path down which the multicast traffic flows from a given multicast source to the last-hop router for troubleshooting purposes.

To do...	Use the command...	Remarks
Trace a multicast path	mtracert <i>source-address</i> [[<i>last-hop-router-address</i>] <i>group-address</i>]	Required Available in any view

Displaying and Maintaining Multicast Routing and Forwarding

To do...	Use the command...	Remarks
View the multicast boundary information	display multicast boundary [<i>group-address</i> [<i>mask</i> <i>mask-length</i>]] [interface <i>interface-type</i> <i>interface-number</i>]	Available in any view
View the multicast forwarding table information	display multicast forwarding-table [<i>source-address</i> [mask { <i>mask</i> <i>mask-length</i> }]] [<i>group-address</i> [mask { <i>mask</i> <i>mask-length</i> }]] incoming-interface { <i>interface-type</i> <i>interface-number</i> register } outgoing-interface { { exclude include match } { <i>interface-type</i> <i>interface-number</i> register } } statistics * [port-info]	Available in any view
View the multicast routing table information	display multicast routing-table [<i>source-address</i> [mask { <i>mask</i> <i>mask-length</i> }]] [<i>group-address</i> [mask { <i>mask</i> <i>mask-length</i> }]] incoming-interface { <i>interface-type</i> <i>interface-number</i> register } outgoing-interface { { exclude include match } { <i>interface-type</i> <i>interface-number</i> register } }] *	Available in any view
View the information of the multicast static routing table	display multicast routing-table static [config] [<i>source-address</i> { <i>mask-length</i> <i>mask</i> }]	Available in any view
View the RPF route information of the specified multicast source	display multicast rpf-info <i>source-address</i> [<i>group-address</i>]	Available in any view
Clear forwarding entries from the multicast forwarding table	reset multicast forwarding-table { { <i>source-address</i> [mask { <i>mask</i> <i>mask-length</i> }]] <i>group-address</i> [mask { <i>mask</i> <i>mask-length</i> }]] incoming-interface { <i>interface-type</i> <i>interface-number</i> register } } * all }	Available in user view
Clear routing entries from the multicast routing table	reset multicast routing-table { { <i>source-address</i> [mask { <i>mask</i> <i>mask-length</i> }]] <i>group-address</i> [mask { <i>mask</i> <i>mask-length</i> }]] incoming-interface { <i>interface-type</i> <i>interface-number</i> register } } * all }	Available in user view



CAUTION:

- The **reset** command clears the information in the multicast routing table or the multicast forwarding table, and thus may cause failure of multicast transmission.
- When a routing entry is deleted from the multicast routing table, the corresponding forwarding entry will also be deleted from the multicast forwarding table.
- When a forwarding entry is deleted from the multicast forwarding table, the corresponding route entry will also be deleted from the multicast routing table.

Configuration Examples

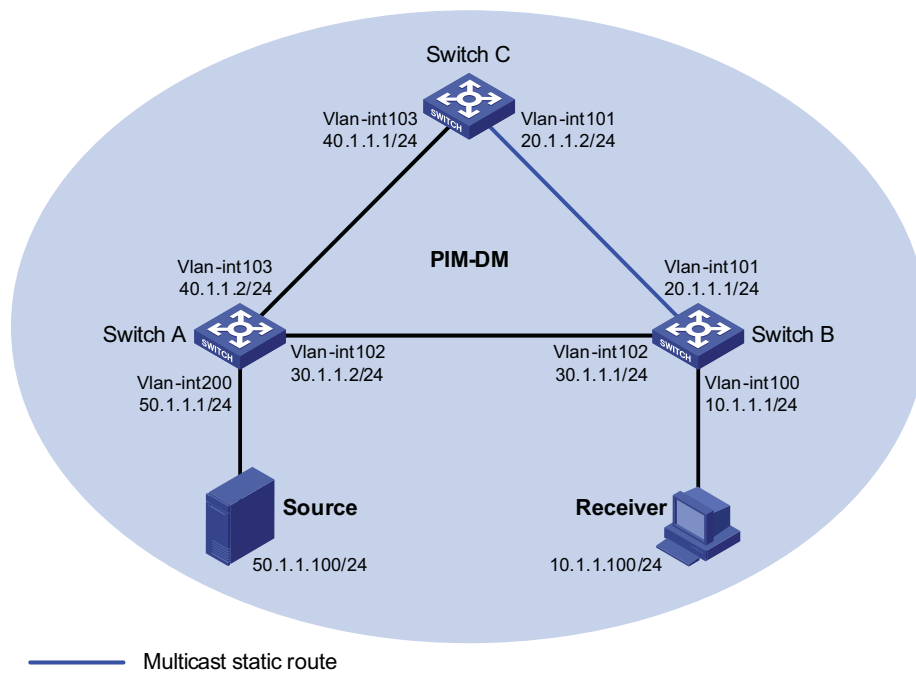
Changing an RPF Route Network requirements

- PIM-DM runs in the network. All switches in the network support multicast.

- Switch A, Switch B and Switch C run OSPF.
- Typically, Receiver can receive the multicast data from Source through the path Switch A - Switch B, which is the same as the unicast route.
- Perform the following configuration so that Receiver can receive the multicast data from Source through the path Switch A - Switch C - Switch B, which is different from the unicast route.

Network diagram

Figure 208 Network diagram for RPF route alternation configuration



Configuration procedure

- 1 Configure the interface IP addresses and enable unicast routing on each switch. Configure the IP address and subnet mask for each interface as per Figure 208. The detailed configuration steps are omitted here.

Enable OSPF on the switches in the PIM-DM domain. Ensure the network-layer interoperation among the switches in the PIM-DM domain. Ensure that the switches can dynamically update their routing information by leveraging the unicast routing protocol. The specific configuration steps are omitted here.

- 2 Enable IP multicast routing, and enable PIM-DM and IGMP
 - # Enable IP multicast routing on Switch B, enable PIM-DM on each interface, and enable IGMP on the host-side interface Ethernet 1/0.

```
<SwitchB> system-view
[SwitchB] multicast routing-enable
[SwitchB] interface vlan-interface 100
[SwitchB-Vlan-interface100] igmp enable
[SwitchB-Vlan-interface100] pim dm
[SwitchB-Vlan-interface100] quit
[SwitchB] interface vlan-interface 101
[SwitchB-Vlan-interface101] pim dm
```

```
[SwitchB-Vlan-interface101] quit
[SwitchB] interface vlan-interface 102
[SwitchB-Vlan-interface102] pim dm
[SwitchB-Vlan-interface102] quit
```

Enable IP multicast routing on Switch A, and enable PIM-DM on each interface.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchA] interface vlan-interface 200
[SwitchA-Vlan-interface200] pim dm
[SwitchA-Vlan-interface200] quit
[SwitchA] interface vlan-interface 102
[SwitchA-Vlan-interface102] pim dm
[SwitchA-Vlan-interface102] quit
[SwitchA] interface vlan-interface 103
[SwitchA-Vlan-interface103] pim dm
[SwitchA-Vlan-interface103] quit
```

The configuration on Switch C is similar to the configuration on Switch A. The specific configuration steps are omitted here.

Use the **display multicast rpf-info** command to view the RPF route to Source on Switch B.

```
[SwitchB] display multicast rpf-info 50.1.1.100
RPF information about source 50.1.1.100:
  RPF interface: Vlan-interface102, RPF neighbor: 30.1.1.2
  Referenced route/mask: 50.1.1.0/24
  Referenced route type: igp
  Route selection rule: preference-preferred
  Load splitting rule: disable
```

As shown above, the current RPF route on Switch B is contributed by a unicast routing protocol and the RPF neighbor is Switch A.

3 Configure a multicast static route

Configure a multicast static route on Switch B, specifying Switch C as its RPF neighbor on the route to Source.

```
[SwitchB] ip rpf-route-static 50.1.1.100 24 20.1.1.2
```

4 Verify the configuration

Use the **display multicast rpf-info** command to view the information about the RPF route to Source on Switch B.

```
[SwitchB] display multicast rpf-info 50.1.1.100
RPF information about source 50.1.1.100:
  RPF interface: Vlan-interface101, RPF neighbor: 20.1.1.2
  Referenced route/mask: 50.1.1.0/24
  Referenced route type: multicast static
  Route selection rule: preference-preferred
  Load splitting rule: disable
```

As shown above, the RPF route on Switch B has changed. It is now the configured multicast static route, and the RPF neighbor is now Switch C.

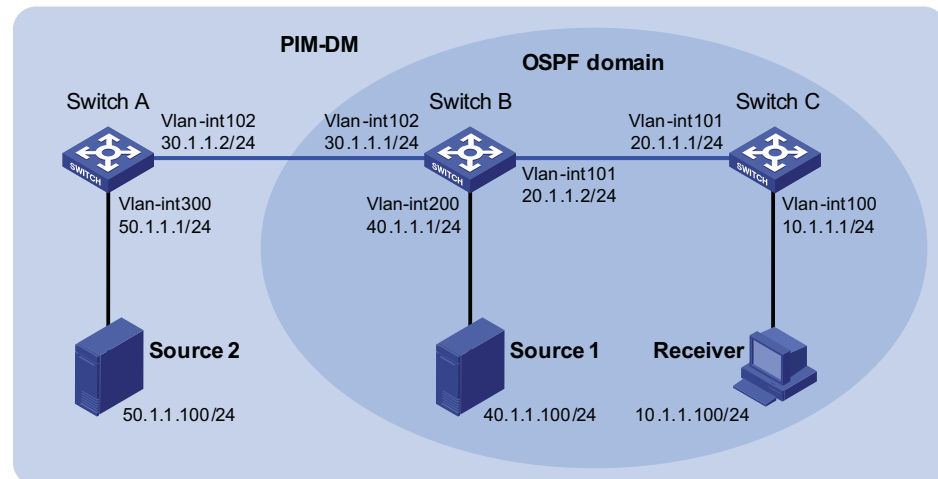
Creating an RPF Route **Network requirements**

- PIM-DM runs in the network and all switches in the network support IP multicast.

- Switch B and Switch C run OSPF, and have no unicast routes to Switch A.
- Typically, Receiver can receive the multicast data from Source 1 in the OSPF domain.
- Perform the following configuration so that Receiver can receive multicast data from Source 2, which is outside the OSPF domain.

Network diagram

Figure 209 Network diagram for creating an RPF route



Configuration procedure

- 1 Configure the interface IP addresses and unicast routing protocol for each switch

Configure the IP address and subnet mask for each interface as per Figure 209. The detailed configuration steps are omitted here.

Enable OSPF on Switch B and Switch C. Ensure the network-layer interoperation among Switch B and Switch C. Ensure that the switches can dynamically update their routing information by leveraging the unicast routing protocol. The specific configuration steps are omitted here.

- 2 Enable IP multicast routing, and enable PIM-DM and IGMP

Enable IP multicast routing on Switch C, enable PIM-DM on each interface, and enable IGMP on the host-side interface VLAN-interface 100.

```
<SwitchC> system-view
[SwitchC] multicast routing-enable
[SwitchC] interface vlan-interface 100
[SwitchC-Vlan-interface100] igmp enable
[SwitchC-Vlan-interface100] pim dm
[SwitchC-Vlan-interface100] quit
[SwitchC] interface vlan-interface 101
[SwitchC-Vlan-interface101] pim dm
[SwitchC-Vlan-interface101] quit
```

Enable IP multicast routing on Switch A and enable PIM-DM on each interface.

```
<SwitchA> system-view
[SwitchA] multicast routing-enable
[SwitchC] interface vlan-interface 300
```

```
[SwitchC-Vlan-interface300] pim dm
[SwitchC-Vlan-interface300] quit
[SwitchC] interface vlan-interface 102
[SwitchC-Vlan-interface102] pim dm
[SwitchC-Vlan-interface102] quit
```

The configuration on Switch B is similar to that on Switch A. The specific configuration steps are omitted here.

Use the **display multicast rpf-info** command to view the RPF routes to Source 2 on Switch B and Switch C.

```
[SwitchB] display multicast rpf-info 50.1.1.100
[SwitchC] display multicast rpf-info 50.1.1.100
```

No information is displayed. This means that no RPF route to Source 2 exists on Switch B and Switch C.

3 Configure a multicast static route

Configure a multicast static route on Switch B, specifying Switch A as its RPF neighbor on the route to Source 2.

```
[SwitchB] ip rpf-route-static 50.1.1.100 24 30.1.1.2
```

Configure a multicast static route on Switch C, specifying Switch B as its RPF neighbor on the route to Source 2.

```
[SwitchC] ip rpf-route-static 50.1.1.100 24 20.1.1.2
```

4 Verify the configuration

Use the **display multicast rpf-info** command to view the RPF routes to Source 2 on Switch B and Switch C.

```
[SwitchB] display multicast rpf-info 50.1.1.100
RPF information about source 50.1.1.100:
  RPF interface: Vlan-interface102, RPF neighbor: 30.1.1.2
  Referenced route/mask: 50.1.1.0/24
  Referenced route type: multicast static
  Route selection rule: preference-preferred
  Load splitting rule: disable
[SwitchC] display multicast rpf-info 50.1.1.100
RPF information about source 50.1.1.100:
  RPF interface: Vlan-interface101, RPF neighbor: 20.1.1.2
  Referenced route/mask: 50.1.1.0/24
  Referenced route type: multicast static
  Route selection rule: preference-preferred
  Load splitting rule: disable
```

As shown above, the RPF routes to Source 2 exist on Switch B and Switch C. The source is the configured static route.

Troubleshooting Multicast Routing and Forwarding

Multicast Static Route Failure

Symptom

No dynamic routing protocol is enabled on the routers, and the physic status and link layer status of interfaces are both up, but the multicast static route fails.

Analysis

- If the multicast static route is not configured or updated correctly to match the current network conditions, the route entry does not exist in the multicast route configuration table and multicast routing table.
- If the optimal route is found, the multicast static route may also fail.

Solution

- 1 In the configuration, you can use the **display multicast routing-table static config** command to view the detailed configuration information of multicast static routes to verify that the multicast static route has been correctly configured and the route entry exists.
- 2 In the configuration, you can use the **display multicast routing-table static** command to view the information of multicast static routes to verify that the multicast static route has been correctly configured and the route entry exists in the multicast routing table.
- 3 Check the next hop interface type of the multicast static route. If the interface is not a point-to-point interface, be sure to specify the next hop address to configure the outgoing interface when you configure the multicast static route.
- 4 Check that the multicast static route matches the specified routing protocol. If a protocol was specified in multicast static route configuration, enter the **display ip routing-table** command to check if an identical route was added by the protocol.
- 5 Check that the multicast static route matches the specified routing policy. If a routing policy was specified when the multicast static route was configured, enter the **display route-policy** command to check the configured routing policy.

Multicast Data Fails to Reach Receivers**Symptom**

The multicast data can reach some routers but fails to reach the last hop router.

Analysis

If a multicast forwarding boundary has been configured through the **multicast boundary** command, any multicast packet will be kept from crossing the boundary.

Solution

- 1 Use the **display pim routing-table** command to check whether the corresponding (S, G) entries exist on the router. If so, the router has received the multicast data; otherwise, the router has not received the data.
- 2 Use the **display multicast boundary** command to view the multicast boundary information on the interfaces. Use the **multicast boundary** command to change the multicast forwarding boundary setting.
- 3 In the case of PIM-SM, use the **display current-configuration** command to check the BSR and RP information.

802.1X CONFIGURATION

When configuring 802.1x, go to these sections for information you are interested in:

- "802.1x Overview" on page 715
- "Configuring 802.1x" on page 726
- "Configuring a Guest VLAN" on page 728
- "Displaying and Maintaining 802.1x" on page 729
- "802.1x Configuration Example" on page 729
- "Guest VLAN Configuration Example" on page 732
- "ACL Assignment Configuration Example" on page 735

802.1x Overview

The 802.1x protocol was proposed by IEEE 802 LAN/WAN committee for security problems on wireless LANs (WLAN). Currently, it is widely used on Ethernet as a common port access control mechanism.

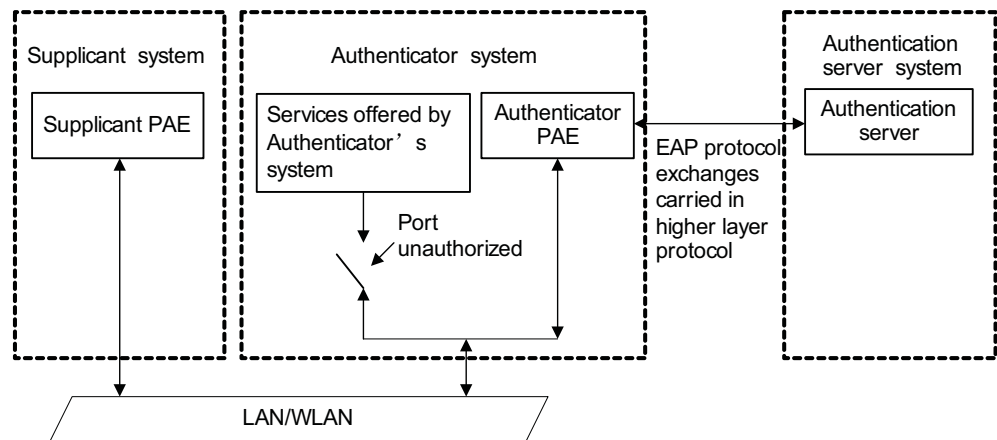
As a port-based network access control protocol, 802.1x authenticates and controls accessing devices at the level of port. A device connected to an 802.1x-enabled port of an access control device can access the resources on the LAN only after passing authentication.

To get more information about 802.1x, go to these topics:

- "Architecture of 802.1x" on page 715
- "Operation of 802.1x" on page 717
- "EAP Encapsulation over LANs" on page 717
- "EAP Encapsulation over RADIUS" on page 719
- "Authentication Process of 802.1x" on page 720
- "802.1x Timers" on page 723
- "Implementation of 802.1x in the Devices" on page 724
- "Features Working Together with 802.1x" on page 724

Architecture of 802.1x

802.1x operates in the typical client/server model and defines three entities: supplicant system, authenticator system, and authentication server system, as shown in Figure 210.

Figure 210 Architecture of 802.1x

- **Supplicant system:** A system at one end of the LAN segment, which is authenticated by the authenticator system at the other end. A supplicant system is usually a user-end device and initiates 802.1x authentication through 802.1x client software supporting the EAP over LANs (EAPOL) protocol.
- **Authenticator system:** A system at the other end of the LAN segment, which authenticates the connected supplicant system. An authenticator system is usually an 802.1x-enabled network device and provides ports (physical or logical) for supplicants to access the LAN.
- **Authentication server system:** The system providing authentication, authorization, and accounting services for the authenticator system. The authentication server, usually a Remote Authentication Dial-in User Service (RADIUS) server, maintains user information like username, password, VLAN that the user belongs to, committed access rate (CAR) parameters, priority, and ACLs.

The above systems involve three basic concepts: PAE, controlled port, control direction.

PAE

Port access entity (PAE) refers to the entity that performs the 802.1x algorithm and protocol operations.

- The authenticator PAE uses the authentication server to authenticate a supplicant trying to access the LAN and controls the status of the controlled port according to the authentication result, putting the controlled port in the state of authorized or unauthorized. In authorized state, the supplicant can access network resources without authentication; in unauthorized state, the supplicant can receive and send EAPOL frames rather than accessing network resources.
- The supplicant PAE responds to the authentication request of the authenticator PAE and provides authentication information. The supplicant PAE can also send authentication requests and logoff requests to the authenticator.

Controlled port and uncontrolled port

An authenticator provides ports for supplicants to access the LAN. Each of the ports can be regarded as two logical ports: a controlled port and an uncontrolled port.

- The uncontrolled port is always open in both the inbound and outbound directions to allow EAPOL protocol frames to pass, guaranteeing that the supplicant can always send and receive authentication frames.
- The controlled port is open to allow normal traffic to pass only when it is in the authorized state.
- The controlled port and uncontrolled port are two parts of the same port. Any frames arriving at the port are visible to both of them.

Control direction

In the unauthorized state, the controlled port can be set to deny traffic to and from the supplicant or just the traffic from the supplicant.

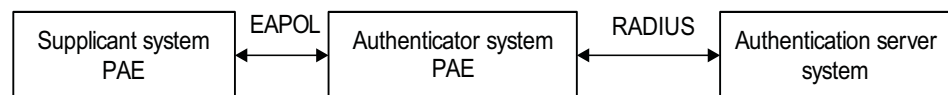


Currently, the devices support only denying the traffic from the supplicant.

Operation of 802.1x

The 802.1x authentication system employs the Extensible Authentication Protocol (EAP) to exchange authentication information between the supplicant PAE, authenticator PAE, and authentication server.

Figure 211 Operation of 802.1x

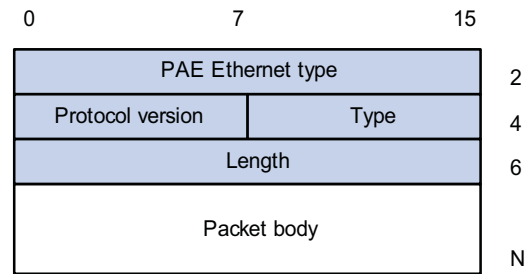


- Between the supplicant PAE and authenticator PAE, EAP protocol packets are encapsulated using EAP Encapsulation over LANs and transferred over the LAN.
- Between the authenticator PAE and authentication server, EAP protocol packets can be handled in two modes: EAP relay and EAP termination. In EAP relay mode, EAP protocol packets are encapsulated by using the EAP Encapsulation over RADIUS (Remote Authentication Dial-In User Service) and then relayed to the RADIUS server. In EAP termination mode, EAP protocol packets are terminated at the authenticator PAE, repackaged in the Password Authentication Protocol (PAP) or Challenge Handshake Authentication Protocol (CHAP) attributes of RADIUS packets, and then transferred to the RADIUS server.
- After a user passes the authentication, the authentication server passes information about the user to the authenticator, which then controls the status of the controlled port according to the instruction of the authentication server.

EAP Encapsulation over LANs

EAPOL frame format

EAPOL, defined by 802.1x, is intended to carry EAP protocol packets between supplicants and authenticators over LANs. Figure 212 shows the EAPOL frame format.

Figure 212 EAPOL frame format

- PAE Ethernet type: Protocol type. It takes the value 0x888E.
- Protocol version: Version of the EAPOL protocol supported by the EAPOL frame sender.
- Type: Type of the EAPOL frame. Table 57 shows the defined types of EAPOL frames.

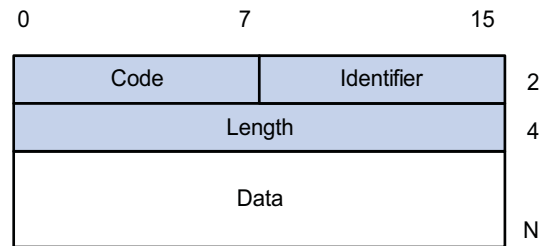
Table 57 Types of EAPOL frames

Type	Description
EAP-Packet (a value of 0x00)	Frame for carrying authentication information, present between an authenticator system and the authentication server. A frame of this type is repackaged and transferred by RADIUS to get through complex networks to reach the authentication server.
EAPOL-Start (a value of 0x01)	Frame for initiating authentication, present between a supplicant and an authenticator.
EAPOL-Logoff (a value of 0x02)	Frame for logoff request, present between a supplicant and an authenticator.
EAPOL-Key (a value of 0x03)	Frame for carrying key information, present between a supplicant and an authenticator.
EAPOL-Encapsulated-ASF-Alert (a value of 0x04)	Frame for carrying alerting information compliant to Alert Standard Forum (ASF). A frame of this type carries network management-related information like warning messages and is terminated at the authenticator.

- Length: Length of the data, that is, length of the Packet body field, in bytes. If the value of this field is 0, no subsequent data field is present.
- Packet body: Content of the packet. The format of this field varies with the value of the Type field.

EAP Packet Format

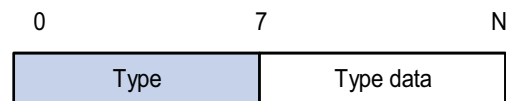
An EAPOL frame of the type of EAP-Packet carries an EAP packet in its Packet body field. The format of the EAP packet is shown in Figure 213.

Figure 213 EAP packet format

- Code: Type of the EAP packet, which can be Request, Response, Success, or Failure.

An EAP packet of the type of Success or Failure has no Data field, and has a length of 4.

An EAP packet of the type of Request or Response has a Data field in the format shown in Figure 214. The Type field indicates the EAP authentication type. A value of 1 represents Identity, indicating that the packet is for querying the identity of the supplicant. A value of 4 represents MD5-Challenge, which corresponds closely to the PPP CHAP protocol.

Figure 214 Format of the Data field in an EAP request/response packet

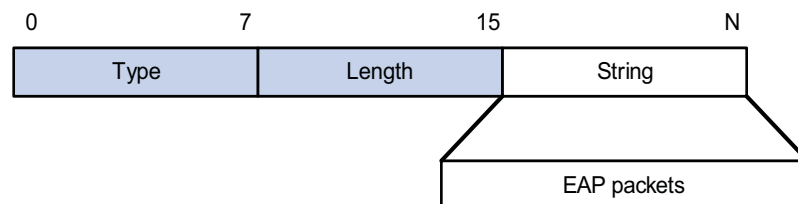
- Identifier: Allows matching of responses with requests.
- Length: Length of the EAP packet, including the Code, Identifier, Length, and Data fields, in bytes.
- Data: Content of the EAP packet. This field is zero or more bytes and its format is determined by the Code field.

EAP Encapsulation over RADIUS

Two attributes of RADIUS are intended for supporting EAP authentication: EAP-Message and Message-Authenticator. For information about RADIUS packet format, refer to "Configuring RADIUS" on page 765.

EAP-Message

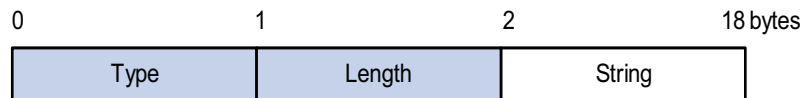
The EAP-Message attribute is used to encapsulate EAP packets. Figure 215 shows its encapsulation format. The value of the Type field is 79. The String field can be up to 253 bytes. If the EAP packet is longer than 253 bytes, it can be fragmented and encapsulated into multiple EAP-Message attributes.

Figure 215 Encapsulation format of the EAP-Message attribute

Message-Authenticator

Figure 216 shows the encapsulation format of the Message-Authenticator attribute. The Message-Authenticator attribute is used to prevent access requests from being snooped during EAP or CHAP authentication. It must be included in any packet with the EAP-Message attribute; otherwise, the packet will be considered invalid and get discarded.

Figure 216 Encapsulation format of the Message-Authenticator attribute



Authentication Process of 802.1x

802.1x authentication can be initiated by either a supplicant or the authenticator system. A supplicant initiates authentication by launching the 802.1x client software to send an EAPOL-Start frame to the authenticator system, while the authenticator system sends an EAP-Request/Identity packet to an unauthenticated supplicant when detecting that the supplicant is trying to login.

An 802.1x authenticator system communicates with a remotely located RADIUS server in two modes: EAP relay and EAP termination. The following description takes the first case as an example to show the 802.1x authentication process.

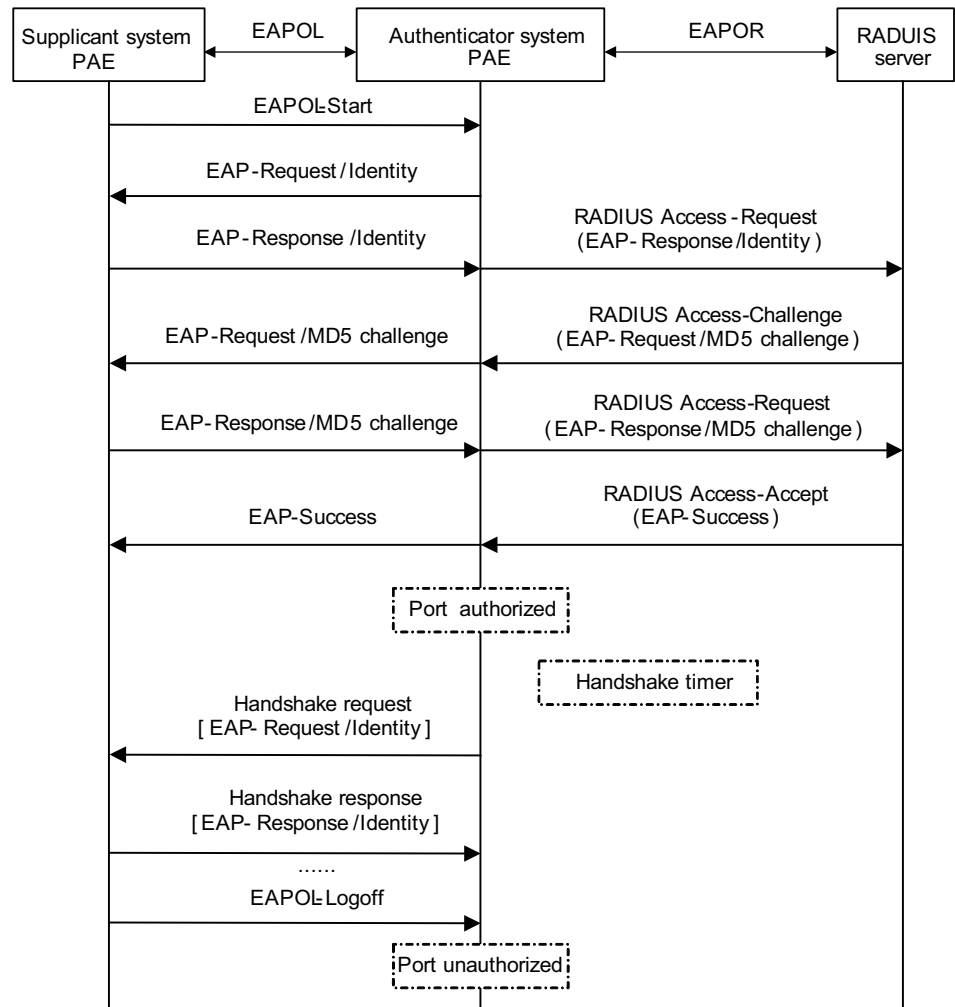
EAP relay

EAP relay is an IEEE 802.1x standard mode. In this mode, EAP packets are carried in an upper layer protocol, such as RADIUS, so that they can go through complex networks and reach the authentication server. Generally, EAP relay requires that the RADIUS server support the EAP attributes of EAP-Message and Message-Authenticator.

At present, the EAP relay mode supports four authentication methods: EAP-MD5, EAP-TLS (Transport Layer Security), EAP-TTLS (Tunneled Transport Layer Security), and PEAP (Protected Extensible Authentication Protocol).

- EAP-MD5: EAP-MD5 authenticates the identity of a supplicant. The RADIUS server sends an MD5 challenge (through an EAP-Request/MD5 Challenge packet) to the supplicant. Then the supplicant encrypts the password with the offered challenge.
- EAP-TLS: With EAP-TLS, a supplicant and the RADIUS server verify each other's security certificates and identities, guaranteeing that EAP packets are sent to the intended destination and thus preventing network traffic from being snooped.
- EAP-TTLS: EAP-TTLS extends EAP-TLS. EAP-TLS allows for mutual authentication between a supplicant and the authentication server. EAP-TTLS extends this implementation by transferring packets through the secure tunnels set up by TLS.
- PEAP: With PEAP, the RADIUS server sets up a TLS tunnel with a supplicant system for integrity protection and then performs a new round of EAP negotiation with the supplicant system for identity authentication.

Figure 217 shows the message exchange procedure with EAP-MD5.

Figure 217 Message exchange in EAP relay mode

- 1 When a user launches the 802.1x client software and enters the registered username and password, the 802.1x client software generates an EAPOL-Start frame and sends it to the authenticator to initiate an authentication process.
- 2 Upon receiving the EAPOL-Start frame, the authenticator responds with an EAP-Request/Identity packet for the username of the supplicant.
- 3 When the supplicant receives the EAP-Request/Identity packet, it encapsulates the username in an EAP-Response/Identity packet and sends the packet to the authenticator.
- 4 Upon receiving the EAP-Response/Identity packet, the authenticator relays the packet in a RADIUS Access-Request packet to the authentication server.
- 5 When receiving the RADIUS Access-Request packet, the RADIUS server compares the identify information against its user information table to obtain the corresponding password information. Then, it encrypts the password information using a randomly generated challenge, and sends the challenge information through a RADIUS Access-Challenge packet to the authenticator.
- 6 After receiving the RADIUS Access-Challenge packet, the authenticator relays the contained EAP-Request/MD5 Challenge packet to the supplicant.

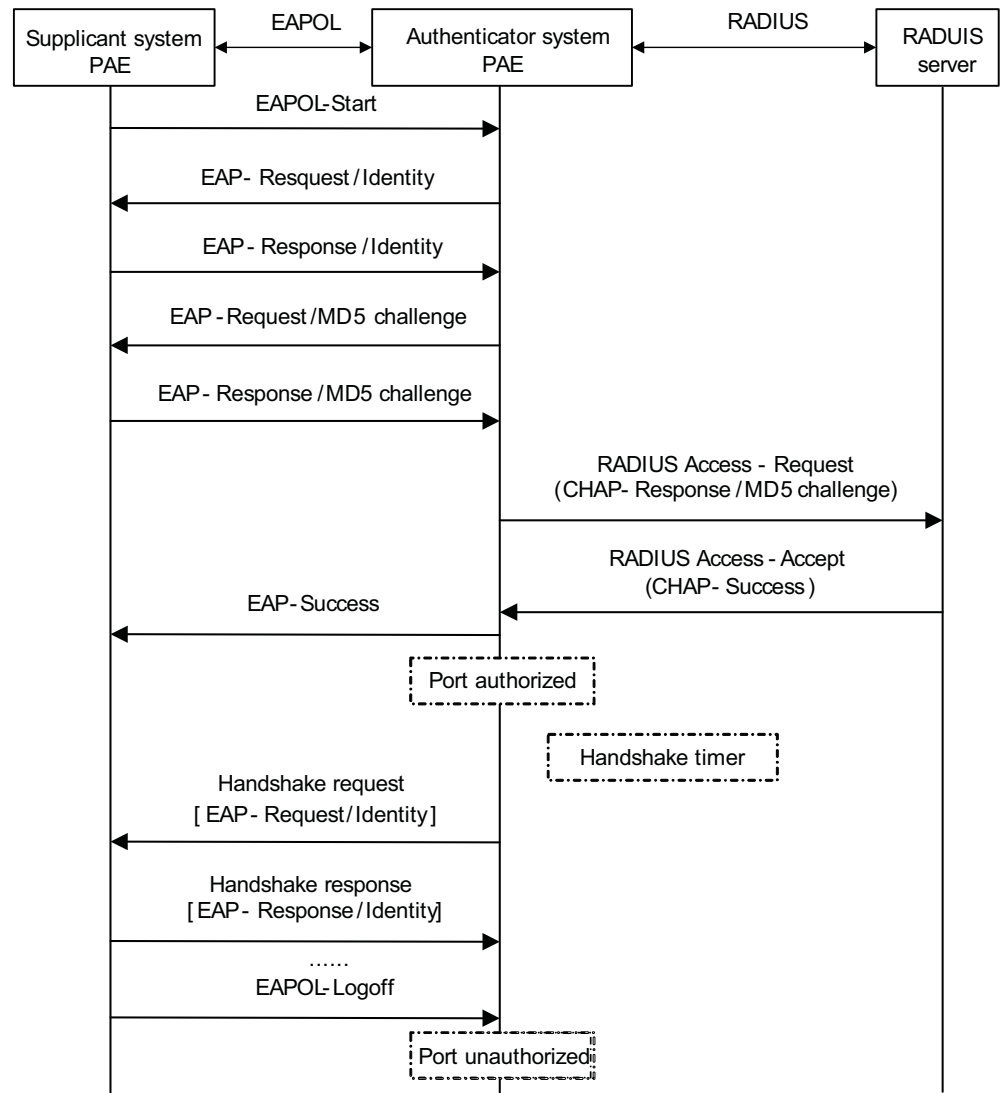
- 7 When receiving the EAP-Request/MD5 Challenge packet, the supplicant uses the offered challenge to encrypt the password part (this process is not reversible), creates an EAP-Response/MD5 Challenge packet, and then sends the packet to the authenticator.
- 8 After receiving the EAP-Response/MD5 Challenge packet, the authenticator relays the packet in a RADIUS Access-Request packet to the authentication server.
- 9 When receiving the RADIUS Access-Request packet, the RADIUS server compares the password information encapsulated in the packet with that generated by itself. If the two are identical, the authentication server considers the user valid and sends to the authenticator a RADIUS Access-Accept packet.
- 10 Upon receiving the RADIUS Access-Accept packet, the authenticator opens the port to grant the access request of the supplicant. After the supplicant gets online, the authenticator periodically sends handshake requests to the supplicant to check whether the supplicant is still online. By default, if two consecutive handshake attempts end up with failure, the authenticator concludes that the supplicant has gone offline and performs the necessary operations, guaranteeing that the authenticator always knows when a supplicant goes offline.
- 11 The supplicant can also send an EAPOL-Logoff frame to the authenticator to go offline unsolicitedly. In this case, the authenticator changes the status of the port from authorized to unauthorized.



*In EAP relay mode, a supplicant must use the same authentication method as that of the RADIUS server, no matter whichever of the above mentioned authentication methods is used. On the device, however, you only need to execute the **dot1x authentication-method eap** command to enable EAP relay.*

EAP termination

In EAP termination mode, EAP packets are terminated at the authenticator and then repackaged into the PAP or CHAP attributes of RADIUS and transferred to the RADIUS server for authentication, authorization, and accounting. Figure 218 shows the message exchange procedure with CHAP authentication.

Figure 218 Message exchange in EAP termination mode

Different from the authentication process in EAP relay mode, it is the authenticator that generates the random challenge for encrypting the user password information in EAP termination authentication process. Consequently, the authenticator sends the challenge together with the username and encrypted password information from the supplicant to the RADIUS server for authentication.

802.1x Timers

Several timers are used in the 802.1x authentication process to guarantee that the supplicants, the authenticators, and the RADIUS server interact with each other in a reasonable manner. The following are the major 802.1x timers:

- Username request timeout timer (tx-period): This timer is used in two cases, one is when an authenticator retransmits an EAP-Request/Identity frame and the other is when an authenticator multicasts an EAP-Request/Identity frame. Once an authenticator sends an EAP-Request/Identity frame to a supplicant, it starts this timer. If this timer expires but it receives no response from the supplicant, it retransmits the request. To cooperate with a supplicant system that does not send EAPOL-Start requests unsolicitedly, the authenticator

multicasts EAP-Request/Identity frames to the supplicant system at an interval defined by this timer.

- Supplicant timeout timer (supp-timeout): Once an authenticator sends an EAP-Request/MD5 Challenge frame to a supplicant, it starts this timer. If this timer expires but it receives no response from the supplicant, it retransmits the request.
- Server timeout timer (server-timeout): Once an authenticator sends a RADIUS Access-Request packet to the authentication server, it starts this timer. If this timer expires but it receives no response from the server, it retransmits the request.
- Handshake timer (handshake-period): After a supplicant passes authentication, the authenticator sends to the supplicant handshake requests at this interval to check whether the supplicant is online. If the authenticator receives no response after sending the allowed maximum number of handshake requests, it considers that the supplicant is offline.
- Quiet timer (quiet-period): When a supplicant fails the authentication, the authenticator refuses further authentication requests from the supplicant in this period of time.

Implementation of 802.1x in the Devices

The devices extend and optimize the mechanism that the 802.1x protocol specifies by:

- Allowing multiple users to access network services through the same physical port.
- Supporting two authentication methods: **portbased** and **macbased**. With the **portbased** method, after the first user of a port passes authentication, all other users of the port can access the network without authentication, and when the first user goes offline, all other users get offline at the same time. With the **macbased** method, each user of a port must be authenticated separately, and when an authenticated user goes offline, no other users are affected.



After an 802.1x supplicant passes authentication, the authentication server sends authorization information to the authenticator. If the authorization information contains VLAN authorization information, the authenticator adds the port connecting the supplicant to the assigned VLAN. This neither changes nor affects the configurations of the port. The only result is that the assigned VLAN takes precedence over the manually configured one, that is, the assigned VLAN takes effect. After the supplicant goes offline, the configured one takes effect.

Features Working Together with 802.1x

VLAN assigning

After an 802.1x user passes the authentication, the server will send an authorization message to the device. If the server is enabled with the VLAN assigning function, the assigned VLAN information will be included in the message. The device, depending on the link type of the port used to log in, adds the port to the assigned VLAN according to the following rules:

- If the port link type is Access, the port leaves its current VLAN and joins the assigned VLAN.
- If the port link type is Trunk, the assigned VLAN is allowed to pass the current trunk port. The default VLAN ID of the port is that of the assigned VLAN.

- If the port link type is Hybrid, the assigned VLAN is allowed to pass the current port without carrying the tag. The default VLAN ID of the port is that of the assigned VLAN.

The assigned VLAN neither changes nor affects the configuration of a port. However, as the assigned VLAN has higher priority than the user-configured VLAN, it is the assigned VLAN that takes effect after a user passes authentication. After the user goes offline, the port returns to its original VLAN.

For details about VLAN configuration, refer to “VLAN Configuration” on page 83.



- *With a Hybrid port, the VLAN assigning will fail if you have configured the assigned VLAN to carry tags.*
- *With a Hybrid port, you cannot configure an assigned VLAN to carry tags after the VLAN has been assigned.*

Guest VLAN

Guest VLAN allows unauthenticated users to access some special resources.

Guest VLAN is the default VLAN that a supplicant on a port can access without authentication. After the supplicant passes 802.1x authentication, the port leaves the guest VLAN and the supplicant can access other network resources.

A user of the guest VLAN can perform operations such as downloading and upgrading the authentication client software. If a supplicant does not have the required authentication client software or the version of the client software is lower, the supplicant will fail the authentication. If no supplicant on a port passes authentication in a certain period of time (45 seconds by default), the port will be added into the guest VLAN.

If a device with 802.1x enabled and the guest VLAN correctly configured sends an EAP-Request/Identity packet for the allowed maximum number of times but gets no response, it adds the port into the guest VLAN according to port link type in the similar way as described in VLAN assigning.

When a supplicant added into the guest VLAN initiates another authentication process, if the authentication is not successful, the supplicant stays in the guest VLAN; otherwise, two cases may occur:

- The authentication server assigns a VLAN: The port leaves the guest VLAN and joins the assigned VLAN. If the supplicant goes offline, the port returns to its original VLAN, that is, the VLAN to which it is configured to belong and it belongs before joining the guest VLAN.
- The authentication server does not assign any VLAN: The port leaves the guest VLAN and returns to its original VLAN. If the supplicant goes offline, the port just stays in its original VLAN.

ACL assignment

ACLs provide a way of controlling access to network resources and defining access rights. When a user logs in through a port, and the RADIUS server is configured with authorization ACLs, the device will permit or deny data flows traversing through the port according to the authorization ACLs. Before specifying authorization ACLs on the server, you need to configure the ACL rules on the

device. You can change the access rights of users by modifying authorization ACL settings on the RADIUS server or changing the corresponding ACL rules on the device.

Configuring 802.1x

Configuration Prerequisites

802.1x provides a user identity authentication scheme. However, 802.1x cannot implement the authentication scheme solely by itself. RADIUS or local authentication must be configured to work with 802.1x.

- Configure the ISP domain to which the 802.1x user belongs and the AAA scheme to be used (that is, local authentication or RADIUS).
- For remote RADIUS authentication, the username and password information must be configured on the RADIUS server.
- For local authentication, the username and password information must be configured on the authenticator and the service type must be set to **lan-access**.

For detailed configuration of the RADIUS client, refer to “Configuring RADIUS” on page 765.

Configuring 802.1x Globally

Follow these steps to configure 802.1x globally:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Enable 802.1x globally		dot1x	Required Disabled by default
Set the authentication method		dot1x authentication-method { chap eap pap }	Optional CHAP by default
Set the port access control parameters	Set the port access control mode for specified or all ports	dot1x port-control { authorized-force auto unauthorized-force } [interface <i>interface-list</i>]	Optional auto by default
	Set the port access control method for specified or all ports	dot1x port-method { macbased portbased } [interface <i>interface-list</i>]	Optional macbased by default
	Set the maximum number of users for specified or all ports	dot1x max-user <i>user-number</i> [interface <i>interface-list</i>]	Optional By default, the maximum number of concurrent users accessing a port is 256.
Set the maximum number of attempts to send an authentication request to a supplicant		dot1x retry <i>max-retry-value</i>	Optional 2 by default

To do...	Use the command...	Remarks
Set timers	dot1x timer { handshake-period <i>handshake-period-value</i> quiet-period <i>quiet-period-value</i> server-timeout <i>server-timeout-value</i> supp-timeout <i>supp-timeout-value</i> tx-period <i>tx-period-value</i> }	Optional The defaults are as follows: 15 seconds for the handshake timer, 60 seconds for the quiet timer, 100 seconds for the server timeout timer, 30 seconds for the supplicant timeout timer, and 30 seconds for the username request timeout timer.
Enable the quiet timer	dot1x quiet-period	Optional Disabled by default



- For 802.1x to take effect on a port, you must enable it both globally in system view and for the port in system view or Ethernet interface view.
- You can also enable 802.1x and set port access control parameters (that is, the port access control mode, port access method, and the maximum number of users) for a port in Ethernet interface view. For detailed configuration, refer to “Configuring 802.1x for a Port” on page 727. The only difference between configuring 802.1x globally and configuring 802.1x for a port lies in the applicable scope. If both a global setting and a local setting exist for an argument of a port, the last configured one is in effect.
- Generally, it is unnecessary to change 802.1x timers unless in some special or extreme network environments.

Configuring 802.1x for a Port

Enabling 802.1x for a port

Follow these steps to enable 802.1x for a port:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable 802.1x for one or more ports	In system view dot1x interface <i>interface-list</i> In Ethernet interface view interface <i>interface-type</i> <i>interface-number</i> dot1x	Required Use either approach. Disabled by default

Configuring 802.1x parameters for a port

Follow these steps to configure 802.1x parameters for a port:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Set the port access control mode for the port	dot1x port-control { authorized-force auto unauthorized-force }	Optional auto by default

To do...	Use the command...	Remarks
Set the port access control method for the port	dot1x port-method { macbased portbased }	Optional macbased by default
Set the maximum number of users for the port	dot1x max-user <i>user-number</i>	Optional By default, the maximum number of concurrent users accessing a port is 256.
Enable online user handshake	dot1x handshake	Optional Enabled by default
Enable multicast trigger	dot1x multicast-trigger	Optional Enabled by default



- You can neither add an 802.1x-enabled port into an aggregation group nor enable 802.1x on a port being a member of an aggregation group.
- Once enabled with the 802.1x multicast trigger function, a port sends multicast trigger messages to the client periodically to initiate authentication.
- For a user-side device sending untagged traffic, the voice VLAN function and 802.1x are mutually exclusive and cannot be configured together on the same port. For details about voice VLAN, refer to “Voice VLAN Configuration” on page 99.
- In EAP relay authentication mode, the authenticator encapsulates the 802.1x user information in the EAP attributes of RADIUS packets and sends the packets to the RADIUS server for authentication. In this case, you can configure the **user-name-format** command but it does not take effect. For information about the **user-name-format** command, refer to “Configuring RADIUS” on page 765.
- If the username of a supplicant contains the version number or one or more blank spaces, you can neither retrieve information nor disconnect the supplicant by using the username. However, you can use items such as IP address and connection index number to do so.

Configuring a Guest VLAN

Configuration Prerequisites

- Enable 802.1x
- Set the port access control method to **portbased** for the port
- Set the port access control mode to **auto** for the port
- Create the VLAN to be specified as the guest VLAN

Configuration Procedure

Follow these steps to configure Guest VLAN:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Configure the guest VLAN for specified or all ports	dot1x guest-vlan <i>vlan-id</i> [interface <i>interface-list</i>] Or in Ethernet interface view interface <i>interface-type</i> <i>interface-number</i> dot1x guest-vlan <i>vlan-id</i>	Required By default, a port is configured with no guest VLAN.



- You can specify a tagged VLAN as the guest VLAN for a Hybrid port, but the guest VLAN does not take effect. Similarly, if a guest VLAN for a Hybrid port is in operation, you cannot configure the guest VLAN to carry tags.
- Configurations in system view are effective to all ports while configurations in interface view are effective to the current port only.
- If a port's access control method is **portbased**, its guest VLAN can take effect; if a port's access control method is **macbased**, its guest VLAN can be configured but cannot take effect.
- A port can be configured with only one guest VLAN. But different ports can have different guest VLANs.



CAUTION: If the data flows from a user-side device include VLAN tags, and 802.1x and guest VLAN are enabled on the access port, you are recommended to configure different VLAN IDs for the Voice VLAN, the default port VLAN, and the guest VLAN of 802.1x.

Displaying and Maintaining 802.1x

To do...	Use the command...	Remarks
Display 802.1x session information, statistics, or configuration information of specified or all ports	display dot1x [sessions statistics] [interface <i>interface-list</i>]	Available in any view
Clear 802.1x statistics	reset dot1x statistics [interface <i>interface-list</i>]	Available in user view

802.1x Configuration Example

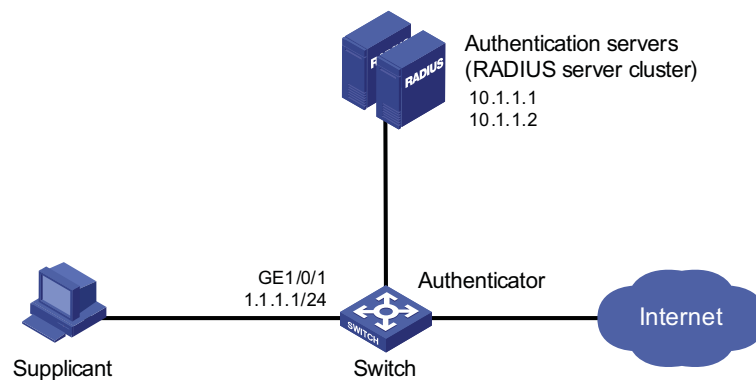
Network requirements

- The access control method of **macbased** is required on the port to control supplicants.
- All supplicants belong to default domain aabbcc.net, which can accommodate up to 30 users. RADIUS authentication is performed at first, and then local authentication when no response from the RADIUS server is received. If the RADIUS accounting fails, the authenticator gets users offline.
- A server group with two RADIUS servers is connected to the switch. The IP addresses of the servers are 10.1.1.1 and 10.1.1.2 respectively. Use the former as the primary authentication/secondary accounting server, and the latter as the secondary authentication/primary accounting server.
- Set the shared key for the switch to exchange packets with the authentication server and the accounting server as secret.

- Specify the switch to try up to five times at an interval of 5 seconds in transmitting a packet to the RADIUS server until it receives a response from the server, and to send real time accounting packets to the accounting server every 15 minutes.
- Specify the switch to remove the domain name from the username before passing the username to the RADIUS server.
- Set the username of the 802.1x user as localuser and the password as localpass and specify to use plain text mode. Enable the idle cut function to get the user offline whenever the user remains idle for over 20 minutes.

Network diagram

Figure 219 Network diagram for 802.1x configuration



Configuration procedure



The following configuration procedure covers most AAA/RADIUS configuration commands for the authenticator, while configuration on the supplicant and RADIUS server are omitted. For information about AAA/RADIUS configuration commands, refer to “Configuring AAA” on page 758 and “Configuring RADIUS” on page 765.

Configure the IP addresses for each interface. (Omitted)

Add local access user localuser, enable the idle cut function, and set the idle cut interval.

```

<Sysname> system-view
[Sysname] local-user localuser
[Sysname-luser-localuser] service-type lan-access
[Sysname-luser-localuser] password simple localpass
[Sysname-luser-localuser] attribute idle-cut 20
[Sysname-luser-localuser] quit
  
```

Create RADIUS scheme radius1 and enter its view.

```

[Sysname] radius scheme radius1
  
```

Configure the IP addresses of the primary authentication and accounting RADIUS servers.

```
[Sysname-radius-radius1] primary authentication 10.1.1.1
[Sysname-radius-radius1] primary accounting 10.1.1.2

# Configure the IP addresses of the secondary authentication and accounting
RADIUS servers.

[Sysname-radius-radius1] secondary authentication 10.1.1.2
[Sysname-radius-radius1] secondary accounting 10.1.1.1

# Specify the shared key for the device to exchange packets with the
authentication server and the accounting server.

[Sysname-radius-radius1] key authentication secret

# Set the interval for the device to retransmit packets to the RADIUS server and the
maximum number of transmission attempts.

[Sysname-radius-radius1] timer response-timeout 5
[Sysname-radius-radius1] retry 5

# Set the interval for the device to send real time accounting packets to the
RADIUS server.

[Sysname-radius-radius1] timer realtime-accounting 15

# Specify the device to remove the domain name of any username before passing
the username to the RADIUS server.

[Sysname-radius-radius1] user-name-format without-domain
[Sysname-radius-radius1] quit

# Create domain aabbcc.net and enter its view.

[Sysname] domain aabbcc.net

# Set radius1 as the RADIUS scheme for users of the domain and specify to use
local authentication as the secondary scheme.

[Sysname-isp-aabbcc.net] authentication default radius-scheme radius
1 local
[Sysname-isp-aabbcc.net] authorization default radius-scheme radius1
local
[Sysname-isp-aabbcc.net] accounting default radius-scheme radius1 lo
cal

# Set the maximum number of users for the domain as 30.

[Sysname-isp-aabbcc.net] access-limit enable 30

# Enable the idle cut function and set the idle cut interval.

[Sysname-isp-aabbcc.net] idle-cut enable 20
[Sysname-isp-aabbcc.net] quit

# Configure aabbcc.net as the default domain.

[Sysname] domain default enable aabbcc.net
```

```
# Enable 802.1x globally.
```

```
[Sysname] dot1x
```

```
# Enable 802.1x for port GigabitEthernet 1/0/1.
```

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] dot1x
[Sysname-GigabitEthernet1/0/1] quit
```

```
# Set the port access control method. (Optional. The default answers the
requirement.)
```

```
[Sysname] dot1x port-method macbased interface GigabitEthernet 1/0/1
```

Guest VLAN Configuration Example

Network requirements

As shown in Figure 220:

- A host is connected to port GigabitEthernet 1/0/1 of the switch and must pass 802.1x authentication to access the Internet.
- The authentication server run RADIUS and is in VLAN 2.
- The update server, which is in VLAN 10, is for client software download and upgrade.
- Port GigabitEthernet 1/0/2 of the switch, which is in VLAN 5, is for accessing the Internet.

As shown in Figure 221:

- On port GigabitEthernet 1/0/1, enable 802.1x and set VLAN 10 as the guest VLAN.

As shown in Figure 222:

- Authenticated supplicants are assigned to VLAN 5 and permitted to access the Internet.

Network diagrams

Figure 220 Network diagram for guest VLAN configuration

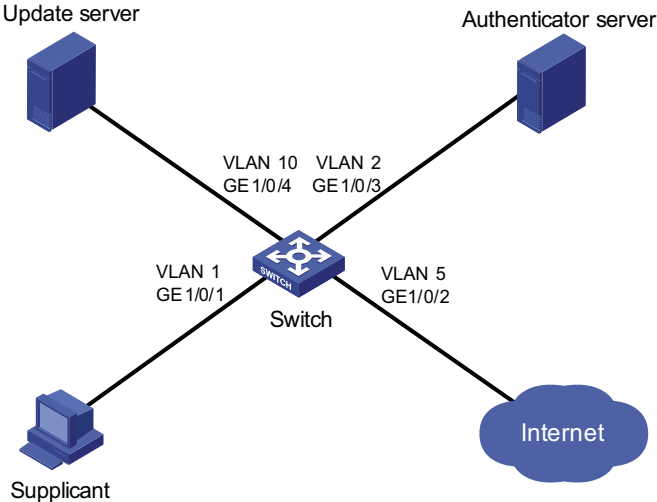


Figure 221 Network diagram with VLAN 10 as the guest VLAN

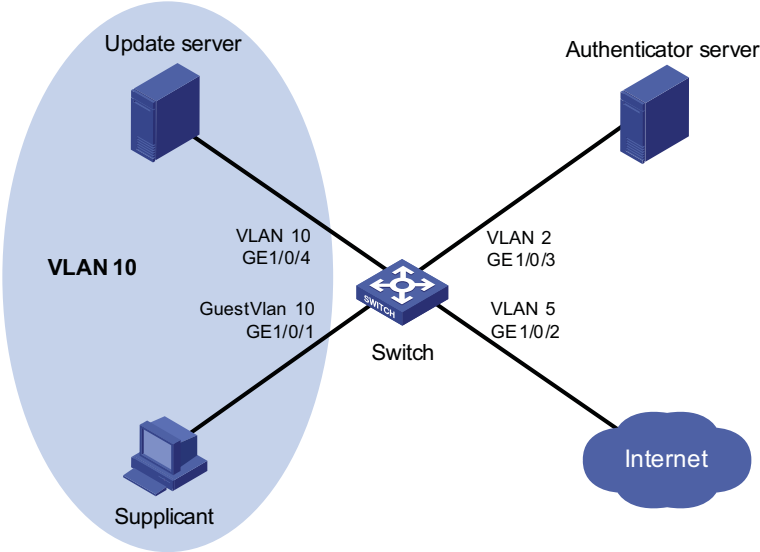
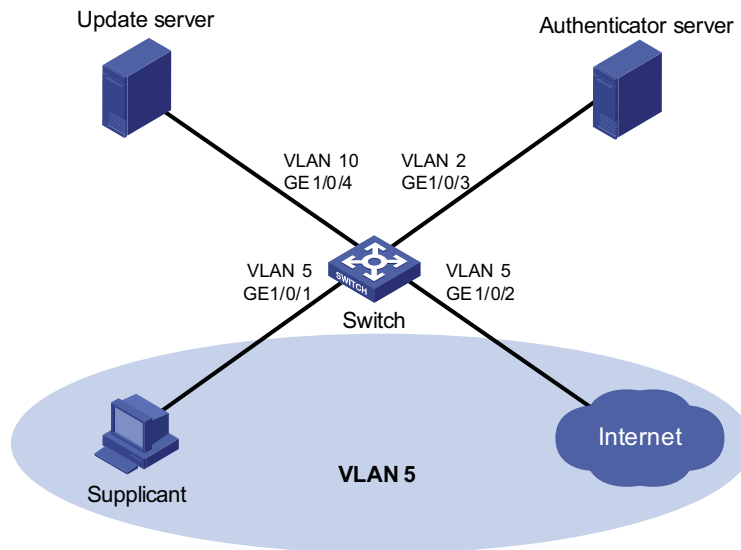


Figure 222 Network diagram when the supplicant passes authentication**Configuration procedure**

Configure RADIUS scheme 2000.

```
<Sysname> system-view
[Sysname] radius scheme 2000
[Sysname-radius-2000] primary authentication 10.11.1.1 1812
[Sysname-radius-2000] primary accounting 10.11.1.1 1813
[Sysname-radius-2000] key authentication abc
[Sysname-radius-2000] key accounting abc
[Sysname-radius-2000] user-name-format without-domain
[Sysname-radius-2000] quit
```

Configure domain system and specify to use RADIUS scheme 2000 for users of the domain.

```
[Sysname] domain system
[Sysname-isp-system] authentication default radius-scheme 2000
[Sysname-isp-system] authorization default radius-scheme 2000
[Sysname-isp-system] accounting default radius-scheme 2000
[Sysname-isp-system] quit
```

Enable 802.1x globally.

```
[Sysname] dot1x
```

Enable 802.1x for port GigabitEthernet 1/0/1.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] dot1x
```

Set the port access control method to **portbased**.

```
[Sysname-GigabitEthernet1/0/1] dot1x port-method portbased
```

Set the port access control mode to **auto**.

```
[Sysname-GigabitEthernet1/0/1] dot1x port-control auto
[Sysname-GigabitEthernet1/0/1] quit
```

Create VLAN 10.

```
[Sysname] vlan 10
[Sysname-vlan10] quit
```

Specify port GigabitEthernet 1/0/1 to use VLAN 10 as its guest VLAN.

```
[Sysname] dot1x guest-vlan 10 interface GigabitEthernet 1/0/1
```

You can use the **display current-configuration** or **display interface GigabitEthernet 1/0/1** command to view your configuration. You can also use the **display vlan 10** command in the following cases to verify whether the configured guest VLAN functions:

- When no users log in.
- When a user fails the authentication.
- When a user goes offline.

ACL Assignment Configuration Example

Network requirements

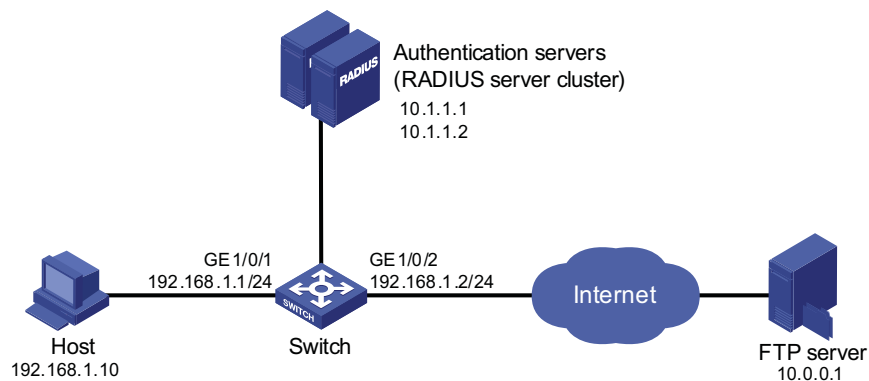
As shown in Figure 223, a host is connected to port GigabitEthernet1/0/1 of the device and must pass 802.1x authentication to access the Internet.

- Configure the RADIUS server to assign ACL 3000.
- Enable 802.1x authentication on GigabitEthernet1/0/1 of the device, and configure ACL 3000.

After the host passes 802.1x authentication, the RADIUS server assigns ACL 3000 to GigabitEthernet1/0/1. As a result, the host can access the Internet but cannot access the FTP server, whose IP address is 10.0.0.1.

Network diagram

Figure 223 Network diagram for ACL assignment



Configuration procedure

```
# Configure the IP addresses of the interfaces. (Omitted)
```

```
# Configure the RADIUS scheme.
```

```

<Sysname> system-view
[Sysname] radius scheme 2000
[Sysname-radius-2000] primary authentication 10.1.1.1 1812
[Sysname-radius-2000] primary accounting 10.1.1.2 1813
[Sysname-radius-2000] key authentication abc
[Sysname-radius-2000] key accounting abc
[Sysname-radius-2000] user-name-format without-domain
[Sysname-radius-2000] quit

```

Create an ISP domain and specify the AAA schemes.

```

[Sysname] domain 2000
[Sysname-isp-2000] authentication default radius-scheme 2000
[Sysname-isp-2000] authorization default radius-scheme 2000
[Sysname-isp-2000] accounting default radius-scheme 2000
[Sysname-isp-2000] quit

```

Configure ACL 3000 to deny packets destined for 10.0.0.1.

```

[Sysname] acl number 3000
[Sysname-acl-adv-3000] rule 0 deny ip destination 10.0.0.1 0

```

Enable 802.1x globally.

```

[Sysname] dot1x

```

Enable 802.1x for GigabitEthernet1/0/1.

```

[Sysname] interface GigabitEthernet1/0/1
[Sysname-GigabitEthernet1/0] dot1x

```

After logging in successfully, a user can use the **ping** command to verify whether the ACL 3000 assigned by the RADIUS server functions.

```

[Sysname] ping 10.0.0.1
PING 10.0.0.1: 56 data bytes, press CTRL_C to break
  Request time out
  Request time out
  Request time out
  Request time out
  Request time out

--- 10.0.0.1 ping statistics ---
  5 packet(s) transmitted
  0 packet(s) received
 100.00% packet loss

```

51

HABP CONFIGURATION

When configuring HABP, go to these sections for the information you are interested in:

- "Introduction to HABP" on page 737
- "Configuring HABP" on page 737
- "Displaying and Maintaining HABP" on page 738

Introduction to HABP

When a switch is configured with the 802.1x function, 802.1x will authenticate and authorize 802.1x-enabled ports and allow only the authorized ports to forward packets. If a port fails 802.1x authentication and authorization, protocol packets passing the port will be blocked. The 3Com Authentication Bypass Protocol (HABP) aims at solving this problem.

On an HABP-capable switch, HABP packets can bypass 802.1x authentication and MAC authentication, allowing communication among switches.

HABP is built on the client-server model. Typically, the HABP server sends HABP requests to the client periodically to collect the MAC address(es) of the attached switch(es). The client responds to the requests, and forwards the HABP requests to the attached switch(es). The HABP server usually runs on the administrative device while the HABP client runs on the attached switches.

Configuring HABP

Complete the following tasks to configure HABP:

- "Configuring the HABP Server" on page 737
- "Configuring an HABP Client" on page 738

Configuring the HABP Server

With enabled with HABP server, the administrative device starts to send HABP requests to the attached switch(es). The HABP responses include the MAC address(es) of the attached switch(es). This makes it possible for the administrative device to manage the attached switch(es).

You only need to configure the interval of sending HABP requests on the administrative device.

Follow these steps to configure an HABP server:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Enable HABP	habp enable	Optional Enabled by default
Configure HABP to work in server mode	habp server vlan <i>vlan-id</i>	Required HABP works in client mode by default.
Set the interval to send HABP requests	habp timer <i>interval</i>	Optional 20 seconds by default

Configuring an HABP Client

Configure HABP to work in client mode on a device connected to the administrative device. Since HABP is enabled and works in client mode by default, this configuration task is optional.

Follow these steps to configure an HABP client:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable HABP	habp enable	Optional Enabled by default
Configure HABP to work in client mode	undo habp server	Optional HABP works in client mode by default.

Displaying and Maintaining HABP

To do...	Use the command...	Remarks
Display HABP configuration information	display habp	Available in any view
Display HABP MAC address table entries	display habp table	Available in any view
Display HABP packet statistics	display habp traffic	Available in any view

MAC AUTHENTICATION CONFIGURATION

When configuring MAC authentication, go to these sections for information you are interested in:

- “MAC Authentication Overview” on page 739
- “Related Concepts” on page 740
- “Configuring MAC Authentication” on page 741
- “Displaying and Maintaining MAC Authentication” on page 742
- “MAC Authentication Configuration Examples” on page 742
- “ACL Assigning Configuration Example” on page 745

MAC Authentication Overview

MAC authentication provides a way for authenticating users based on ports and MAC addresses, without requiring any client software to be installed on the hosts. Once detecting a new MAC address, it initiates the authentication process without requiring username or password.

Currently, the device supports two MAC authentication modes:

- Remote Authentication Dial-In User Service (RADIUS) based MAC authentication
- Local MAC authentication

For detailed information about RADIUS authentication and local authentication, refer to “Configuring RADIUS” on page 765.

After determining the authentication mode to be used, you can choose the type of MAC authentication username, including:

- MAC address, where the MAC address of a user serves as both the username and password.
- Fixed username, where all users use the same preconfigured username and password for authentication, regardless of the MAC addresses.

RADIUS-Based MAC Authentication

In RADIUS-base MAC authentication, the device serves as a RADIUS client and requires a RADIUS server to cooperate with it.

- If the type of MAC authentication username is MAC address, the device forwards a detected MAC address as the username and password to the RADIUS server for authentication of the user.
- If the type of MAC authentication username is fixed username, the device sends the same username and password configured locally to the RADIUS server for authentication of each user.

If the authentication succeeds, the user will be granted permission to access the network resources.

Local MAC Authentication

In local MAC authentication, the device performs authentication of users locally and different items need to be manually configured for users on the device according to the type of MAC authentication username:

- If the type of MAC authentication username is MAC address, a local user must be configured for each user on the device, using the MAC address of the user as both the username and password.
- If the type of MAC authentication username is fixed username, a single username and optionally a single password are required for the device to authenticate all users.

Related Concepts

MAC Authentication Timers

The following timers function in the process of MAC authentication:

- Offline detect timer: At this interval, the device checks to see whether an online user has gone offline. Once detecting that a user becomes offline, the device sends to the RADIUS server a stop accounting notice.
- Quiet timer: Whenever a user fails MAC authentication, the device does not initiate any MAC authentication of the user during such a period.
- Server timeout timer: During authentication of a user, if the device receives no response from the RADIUS server in this period, it assumes that its connection to the RADIUS server has timed out and forbids the user from accessing the network.

Quiet MAC Address

When a user fails MAC authentication, the MAC address becomes a quiet MAC address, which means that any packets from the MAC address will be discarded simply by the device until the quiet timer expires. This prevents the device from authenticating invalid users repeatedly in a short time.



CAUTION: *If the quiet MAC is the same as the static MAC configured or an authentication-passed MAC, then the quiet function is not effective.*

VLAN Assigning

For separation of users from restricted network resources, a more general way is to put the users and restricted resources into different VLANs. After a user passes identity authentication, the authorization server assigns the VLAN where the restricted resources reside as an authorized VLAN and the port to which the user is connected will become a member of the authorized VLAN. As a result, the user can access those restricted network resources.

ACL Assigning

ACLs assigned by an authorization server are referred to as authorization ACLs, which are designed to control access to network resources with a very fine granularity. When a user logs in, if the RADIUS server is configured with authorization ACLs, the device will permit or deny data flows traversing through the port through which the user accesses the device according to the authorization ACLs assigned by the RADIUS server. You can change access rights of users by modifying authorization ACL settings on the RADIUS server.

Configuring MAC Authentication

Configuration Prerequisites

- Create and configure an ISP domain.
- For local authentication, create the local users and configure the passwords.
- For RADIUS authentication, ensure that a route is available between the device and the RADIUS server.



CAUTION: For local authentication

- The type of username and password of a local user must be consistent with that used for MAC authentication.
- All the letters in the MAC address to be used as the username and password of a local user must be in lower case.
- The service type of the local user must be configured as **lan-access**.

Configuration Procedure

Follow these steps to configure MAC authentication:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable MAC authentication globally	mac-authentication	Required Disabled by default
Enable MAC authentication for specified ports	mac-authentication interface <i>interface-list</i> interface <i>interface-type</i> <i>interface-number</i> mac-authentication quit	Required Disabled by default
Specify the ISP domain for MAC authentication	mac-authentication domain <i>isp-name</i>	Optional The default ISP domain (system) is used by default.
Set the offline detect timer	mac-authentication timer offline-detect <i>offline-detect-value</i>	Optional 300 seconds by default
Set the quiet timer	mac-authentication timer quiet <i>quiet-value</i>	Optional 60 seconds by default
Set the server timeout timer	mac-authentication timer server-timeout <i>server-timeout-value</i>	Optional 100 seconds by default
Configure the username and password for MAC authentication	mac-authentication user-name-format { fixed [account <i>name</i>] [password { cipher simple } <i>password</i>] mac-address [with-hyphen without-hyphen] }	Optional By default, the user's source MAC address serves as the username and password, and the MAC address does not contain hyphen "-".



- You can configure MAC authentication for various ports in advance. The configuration, however, takes effect only after the global MAC authentication is enabled.

- You can neither add a MAC authentication enabled port into an aggregation group, nor enable MAC authentication on a port added into an aggregation group.

Displaying and Maintaining MAC Authentication

To do...	Use the command...	Remarks
Display the global MAC authentication information or the MAC authentication information about specified ports	display mac-authentication [interface <i>interface-list</i>]	Available in any view
Clear the MAC authentication statistics	reset mac-authentication statistics [interface <i>interface-list</i>]	Available in user view

MAC Authentication Configuration Examples

Local MAC Authentication Configuration Example

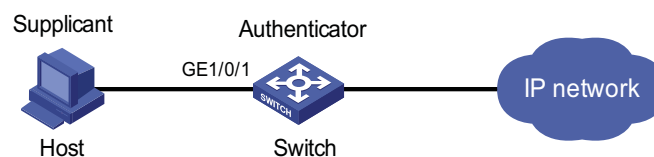
Network requirements

As illustrated in Figure 224, a supplicant is connected to the device through port GigabitEthernet 1/0/1.

- Local MAC authentication is required on every port to control user access to the Internet.
- All users belong to domain **aabbcc.net**.
- A local user uses **aaa** as the username and **123456** as the password for authentication.
- Set the offline detect timer to 180 seconds and the quiet timer to 3 minutes.

Network Diagram

Figure 224 Network diagram for local MAC authentication



Configuration Procedure

- 1 Configure MAC authentication on the switch.

Add a local user.

```

<Sysname> system-view
[Sysname] local-user aaa
[Sysname-luser-aaa] password simple 123456
[Sysname-luser-aaa] service-type lan-access
[Sysname-luser-aaa] quit
  
```

Configure ISP domain **aabbcc.net**, and specify to perform local authentication.

```

[Sysname] domain aabbcc.net
[Sysname-isp-aabbcc.net] authentication lan-access local
[Sysname-isp-aabbcc.net] quit

# Enable MAC authentication globally.

[Sysname] mac-authentication

# Enable MAC authentication for port GigabitEthernet 1/0/1.

[Sysname] mac-authentication interface GigabitEthernet 1/0/1

# Specify the ISP domain for MAC authentication.

[Sysname] mac-authentication domain aabbcc.net

# Set the MAC authentication timers.

[Sysname] mac-authentication timer offline-detect 180
[Sysname] mac-authentication timer quiet 3
[Sysname] mac-authentication user-name-format fixed account aaa password simple 123456

```

1 Verify the configuration

Display global MAC authentication information.

```

<Sysname> display mac-authentication
MAC address authentication is Enabled.
User name format is fixed account
  Fixed username:aaa
  Fixed password:123456
    Offline detect period is 180s
    Quiet period is 60s.
    Server response timeout value is 100s
    The max allowed user number is 1024 per slot
    Current user number amounts to 1
    Current domain is aabbcc.net
Silent Mac User info:
  MAC ADDR          From Port          Port Index
GigabitGigabitEthernet1/0/1 is link-up
  MAC address authentication is Enabled
  Authenticate success: 1, failed: 0
  Current online user number is 1
  MAC ADDR          Authenticate state          AuthIndex
  00e0-fc12-3456    MAC_AUTHENTICATOR_SUCCESS    29

```

RADIUS-Based MAC Authentication Configuration Example

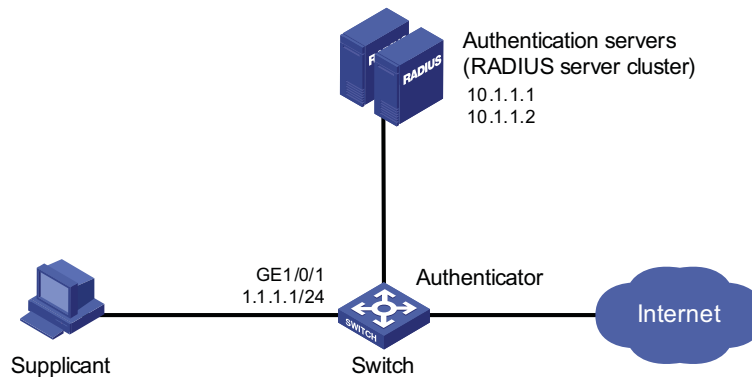
Network requirements

As illustrated in Figure 225, a host is connected to the device through port GigabitEthernet 1/0/1. The device authenticates the host through the RADIUS server.

- MAC authentication is required on every port to control user access to the Internet.
- Set the offline detect timer to 180 seconds and the quiet timer to 3 minutes.

Network diagram

Figure 225 Network diagram for MAC authentication using RADIUS



Configuration procedure

1 Configure MAC authentication on the device

Configure the IP addresses of the interfaces. (Omitted)

Configure a RADIUS scheme.

```
<Sysname> system-view
[Sysname] radius scheme 2000
[Sysname-radius-2000] primary authentication 10.1.1.1 1812
[Sysname-radius-2000] primary accounting 10.1.1.2 1813
[Sysname-radius-2000] key authentication abc
[Sysname-radius-2000] key accounting abc
[Sysname-radius-2000] user-name-format without-domain
[Sysname-radius-2000] quit
```

Specify the AAA schemes for the ISP domain.

```
[Sysname] domain 2000
[Sysname-isp-2000] authentication default radius-scheme 2000
[Sysname-isp-2000] authorization default radius-scheme 2000
[Sysname-isp-2000] accounting default radius-scheme 2000
[Sysname-isp-2000] quit
```

Enable MAC authentication globally.

```
[Sysname] mac-authentication
```

Enable MAC authentication for port GigabitEthernet 1/0/1.

```
[Sysname] mac-authentication interface GigabitEthernet 1/0/1
```

Specify the ISP domain for MAC authentication.

```
[Sysname] mac-authentication domain 2000
```

Set the MAC authentication timers.

```
[Sysname] mac-authentication timer offline-detect 180
[Sysname] mac-authentication timer quiet 3
```

```
[Sysname] mac-authentication user-name-format fixed account aaa pass
word simple 123456
```

2 Verify the configuration

Display global MAC authentication information.

```
<Sysname> display mac-authentication
MAC address authentication is Enabled.
User name format is fixed account
Fixed username:aaa
Fixed password:123456
Offline detect period is 180s
Quiet period is 60s.
Server response timeout value is 100s
The max allowed user number is 1024 per slot
Current user number amounts to 1
Current domain is 2000
Silent Mac User info:
      MAC ADDR                From Port                Port Index
GigabitGigabitEthernet1/0/1 is link-up
MAC address authentication is Enabled
Authenticate success: 1, failed: 0
Current online user number is 1
      MAC ADDR                Authenticate state        AuthIndex
00e0-fc12-3456  MAC_AUTHENTICATOR_SUCCESS  29
```

ACL Assigning Configuration Example

Network requirements

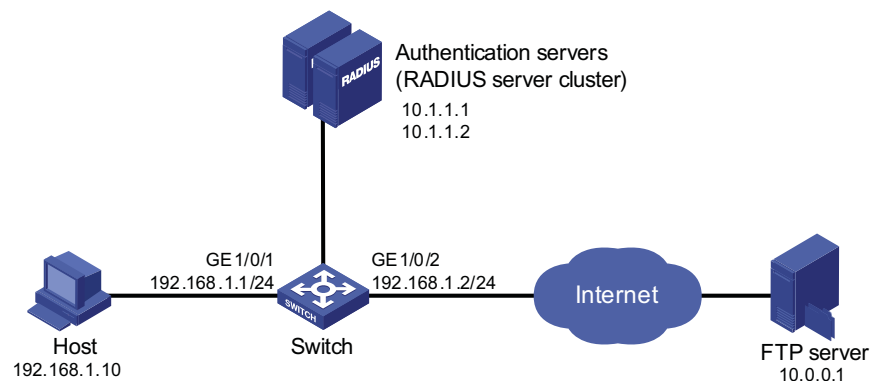
As shown in Figure 226, a host is connected to port GigabitEthernet1/0/1 of the switch and must pass MAC authentication to access the Internet.

- Configure the RADIUS server to assign ACL 3000.
- On port Ethernet 1/0 of the switch, enable MAC authentication and configure ACL 3000.

After the host passes MAC authentication, the RADIUS server assigns ACL 3000 to port Ethernet 1/0 of the switch. As a result, the host can access the Internet but cannot access the FTP server, whose IP address is 10.0.0.1.

Network diagram

Figure 226 Network diagram for ACL assigning



Configuration procedure

Configure the IP addresses of the interfaces. (Omitted)

Configure the RADIUS scheme.

```
<Sysname> system-view
[Sysname] radius scheme 2000
[Sysname-radius-2000] primary authentication 10.1.1.1 1812
[Sysname-radius-2000] primary accounting 10.1.1.2 1813
[Sysname-radius-2000] key authentication abc
[Sysname-radius-2000] key accounting abc
[Sysname-radius-2000] user-name-format without-domain
[Sysname-radius-2000] quit
```

Create an ISP domain and specify the AAA schemes.

```
[Sysname] domain 2000
[Sysname-isp-2000] authentication default radius-scheme 2000
[Sysname-isp-2000] authorization default radius-scheme 2000
[Sysname-isp-2000] accounting default radius-scheme 2000
[Sysname-isp-2000] quit
```

Configure ACL 3000 to deny packets destined for 10.0.0.1.

```
[Sysname] acl number 3000
[Sysname-acl-adv-3000] rule 0 deny ip destination 10.0.0.1 0
[Sysname-acl-adv-3000] quit
```

Enable MAC authentication globally.

```
[Sysname] mac-authentication
```

Enable MAC authentication for port GigabitEthernet1/0/1.

```
[Sysname] interface GigabitEthernet 1/0/1.
[Sysname- GigabitEthernet1/0/1] mac-authentication
```

After completing the above configurations, you can use the **ping** command to verify whether the ACL 3000 assigned by the RADIUS server functions.

```
[Sysname] ping 10.0.0.1
PING 10.0.0.1: 56 data bytes, press CTRL_C to break
Request time out
Request time out
Request time out
Request time out
Request time out

--- 10.0.0.1 ping statistics ---
5 packet(s) transmitted
0 packet(s) received
100.00% packet loss
```

AAA/RADIUS/HWTACACS CONFIGURATION

When configuring AAA/RADIUS/HWTACACS, go to these sections for information you are interested in:

- “AAA/RADIUS/HWTACACS Overview” on page 747
- “AAA/RADIUS/HWTACACS Configuration Task List” on page 756
- “Configuring AAA” on page 758
- “Configuring RADIUS” on page 765
- “Configuring HWTACACS” on page 771
- “Displaying and Maintaining AAA/RADIUS/HWTACACS” on page 775
- “AAA/RADIUS/HWTACACS Configuration Examples” on page 776
- “Troubleshooting AAA/RADIUS/HWTACACS” on page 779

AAA/RADIUS/HWTACACS Overview

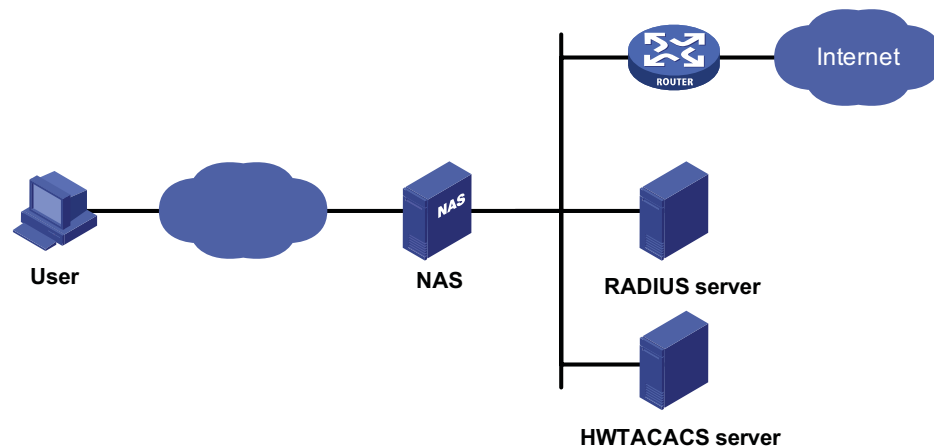
This section covers these topics:

- “Introduction to AAA” on page 747
- “Introduction to RADIUS” on page 749
- “Introduction to HWTACACS” on page 754

Introduction to AAA

Authentication, Authorization, and Accounting (AAA) provides a uniform framework for configuring these three security functions to implement network security management.

AAA usually uses a client/server model, where the client runs on the network access server (NAS) and the server maintains user information centrally. In an AAA network, a NAS is a server for users but a client for the AAA servers, as shown in Figure 227.

Figure 227 AAA networking diagram

When a user tries to establish a connection to the NAS and obtain the rights to access other networks or some network resources, the NAS authenticates the user or the corresponding connection. The NAS can also transparently pass the user authentication, authorization and accounting information to the server (RADIUS server or HWTACACS server). The RADIUS/HWTACACS protocol defines how to exchange user information between a NAS and a server.

In the AAA network shown in Figure 227, there is a RADIUS server and a HWTACACS server. You can determine the authentication, authorization and accounting scheme according to the actual requirements. For example, you can use the RADIUS server for authentication and authorization, and the HWTACACS server for accounting.

The three security functions are described as follows:

- **Authentication:** Identifies remote users and judges whether a user is legal.
- **Authorization:** Grants different users different rights. For example, a user logging into the server can be granted the permission to access and print the files in the server.
- **Accounting:** Records all network service usage information of users, including the service type, start and end time, and traffic. In this way, accounting can be used for not only accounting itself, but also network security surveillance.

You can use AAA to provide only one or two security functions, if desired. For example, if your company only wants employees to be authenticated before they access specific resources, you can configure only an authentication server. If the network usage information is expected to be recorded, you also need to configure an accounting server.

As mentioned above, AAA provides a uniform framework to implement network security management. It is a security mechanism that enables authenticated and authorized entities to access specific resources and records operations by the entities. The AAA framework thus allows for excellent scalability and centralized user information management.

AAA can be implemented through multiple protocols. Currently, the device supports using RADIUS and HWTACACS for AAA, and RADIUS is often used in practice.

Introduction to RADIUS

Remote Authentication Dial-In User Service (RADIUS) is a distributed information interaction protocol in the client/server model. RADIUS can protect networks against unauthorized access and is often used in network environments where both high security and remote user access are required. Based on UDP, RADIUS defines the RADIUS packet format and the message transfer mechanism, and uses UDP port 1812 as the authentication port and 1813 as the accounting port.

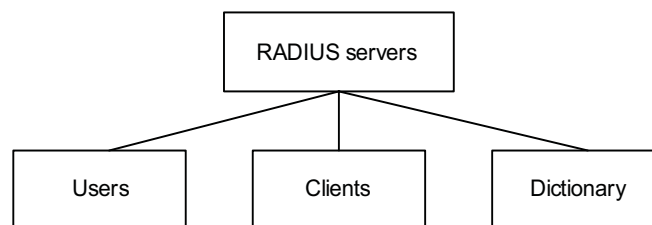
RADIUS was originally designed for dial-in user access. With the diversification of access methods, RADIUS has been extended to support more access methods, for example, Ethernet access and ADSL access. It uses authentication and authorization to provide access service and uses accounting to collect and record usage of network resources by users.

Client/server model

- Client: The RADIUS client runs on the NASs located throughout the network. It passes user information to designated RADIUS servers and acts on the response (for example, rejects or accepts user access requests).
- Server: The RADIUS server runs on the computer or workstation at the network center and maintains information related to user authentication and network service access. It authenticates a user after receiving a connection request and returns the processing result (for example, rejecting or accepting user access requests) to the client.

In general, the RADIUS server maintains three databases, namely, Users, Clients, and Dictionary, as shown in Figure 228:

Figure 228 RADIUS server components



- Users: Stores user information such as the username, password, applied protocols, and IP address.
- Clients: Stores information about RADIUS clients such as the shared keys and IP addresses.
- Dictionary: Stores the information for interpreting RADIUS protocol attributes and their values.

Security authentication mechanism

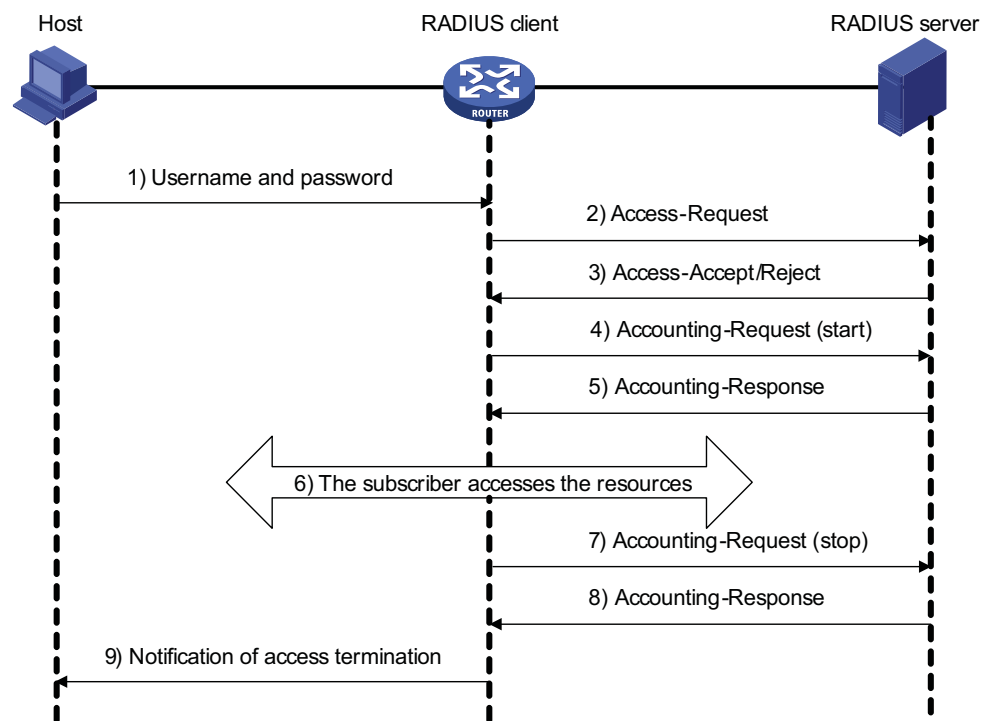
Information exchanged between the RADIUS client and the RADIUS server is authenticated with a shared key, which is never transmitted over the network, thus enhancing the security of information exchange. To prevent user passwords from being intercepted in non-secure networks, the passwords are encrypted during transmission.

A RADIUS server supports multiple user authentication methods, such as the Password Authentication Protocol (PAP) and Challenge Handshake Authentication Protocol (CHAP) of Point-to-Point Protocol (PPP). In addition, a RADIUS server can act as the client of another AAA server to provide proxy authentication or accounting service.

Basic message exchange process of RADIUS

For the interaction among the host, the RADIUS client, and the RADIUS server, see Figure 229.

Figure 229 Basic message exchange process of RADIUS



The following is how RADIUS operates:

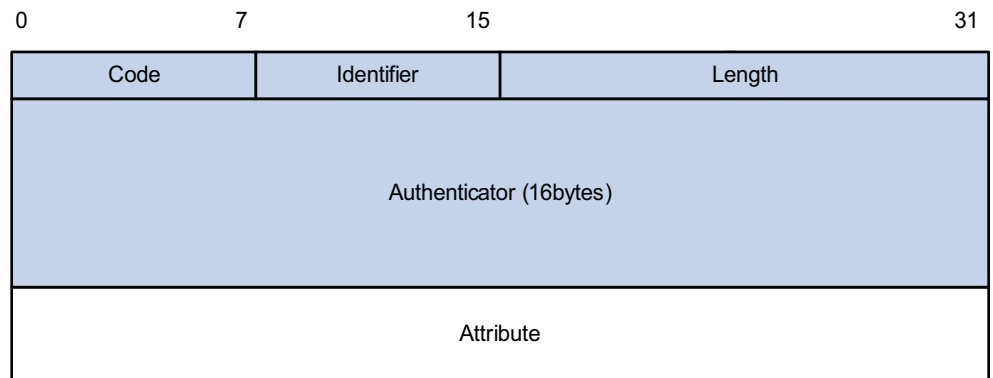
- 1** The host initiates a connection request carrying the username and password to the RADIUS client.
- 2** Having received the username and password, the RADIUS client sends an authentication request (Access-Request) to the RADIUS server, where the user password is encrypted by the Message-Digest 5 (MD5) algorithm with the shared key.
- 3** The RADIUS server authenticates the username and password. If the authentication succeeds, it sends back an Access-Accept message containing the information of user's right. If the authentication fails, it returns an Access-Reject message.
- 4** The RADIUS client accepts or denies the user according to the returned authentication result. If it accepts the user, it sends a start-accounting request (Accounting-Request) to the RADIUS server.
- 5** The RADIUS server returns a start-accounting response (Accounting-Response) and starts accounting.

- 6 The subscriber accesses the network resources.
- 7 The host requests the RADIUS client to tear down the connection and the RADIUS client sends a stop-accounting request (Accounting-Request) to the RADIUS server.
- 8 The RADIUS server returns a stop-accounting response (Accounting-Response) and stops accounting.
- 9 The subscriber stops network resource accessing.

RADIUS packet structure

RADIUS uses UDP to transmit messages. It ensures the smooth message exchange between the RADIUS server and the client through a series of mechanisms, including the timer management mechanism, retransmission mechanism, and slave server mechanism. Figure 230 shows the RADIUS packet structure.

Figure 230 RADIUS packet structure



Descriptions of fields are as follows:

- 1 The Code field (1-byte long) is for indicating the type of the RADIUS packet. Table 58 gives the possible values and their meanings.

Table 58 Main values of the Code field

Code	Packet type	Description
1	Access-Request	From the client to the server. A packet of this type carries user information for the server to authenticate the user. It must contain the User-Name attribute and can optionally contain the attributes of NAS-IP-Address, User-Password, and NAS-Port.
2	Access-Accept	From the server to the client. If all the attribute values carried in the Access-Request are acceptable, that is, the authentication succeeds, the server sends an Access-Accept response.
3	Access-Reject	From the server to the client. If any attribute value carried in the Access-Request is unacceptable, the server rejects the user and sends an Access-Reject response.
4	Accounting-Request	From the client to the server. A packet of this type carries user information for the server to start/stop accounting on the user. It contains the Acct-Status-Type attribute, which indicates whether the server is requested to start the accounting or to end the accounting.

Table 58 Main values of the Code field

Code	Packet type	Description
5	Accounting-Response	From the server to the client. The server sends to the client a packet of this type to notify that it has received the Accounting-Request and has correctly recorded the accounting information.

- 2 The Identifier field (1-byte long) is for matching request packets and response packets and detecting retransmitted request packets. The request and response packets of the same type have the same identifier.
- 3 The Length field (2-byte long) indicates the length of the entire packet, including the Code, Identifier, Length, Authenticator, and Attribute fields. The value of the field is in the range 20 to 4096. Bytes beyond the length are considered the padding and are neglected after being received. If the length of a received packet is less than that indicated by the Length field, the packet is dropped.
- 4 The Authenticator field (16-byte long) is used to authenticate the reply from the RADIUS server, and is also used in the password hiding algorithm. There are two kinds of authenticators: Request authenticator and Response authenticator.
- 5 The Attribute field carries information about the configuration details of a request or response. This field is represented in triplets of Type, Length, and Value.
- 6 Type: One byte, in the range 1 to 255. It indicates the type of the attribute. Commonly used attributes for RADIUS authentication and authorization are listed in Table 59.
- 7 Length: One byte for indicating the length of the attribute in bytes, including the Type, Length, and Value fields.
- 8 Value: Value of the attribute, up to 253 bytes. Its format and content depend on the Type and Length fields.

Table 59 RADIUS attributes

No.	Attribute type	No.	Attribute type
1	User-Name	45	Acct-Authentic
2	User-Password	46	Acct-Session-Time
3	CHAP-Password	47	Acct-Input-Packets
4	NAS-IP-Address	48	Acct-Output-Packets
5	NAS-Port	49	Acct-Terminate-Cause
6	Service-Type	50	Acct-Multi-Session-Id
7	Framed-Protocol	51	Acct-Link-Count
8	Framed-IP-Address	52	Acct-Input-Gigawords
9	Framed-IP-Netmask	53	Acct-Output-Gigawords
10	Framed-Routing	54	(unassigned)
11	Filter-ID	55	Event-Timestamp
12	Framed-MTU	56-59	(unassigned)
13	Framed-Compression	60	CHAP-Challenge
14	Login-IP-Host	61	NAS-Port-Type
15	Login-Service	62	Port-Limit
16	Login-TCP-Port	63	Login-LAT-Port
17	(unassigned)	64	Tunnel-Type

Table 59 RADIUS attributes

No.	Attribute type	No.	Attribute type
18	Reply_Message	65	Tunnel-Medium-Type
19	Callback-Number	66	Tunnel-Client-Endpoint
20	Callback-ID	67	Tunnel-Server-Endpoint
21	(unassigned)	68	Acct-Tunnel-Connection
22	Framed-Route	69	Tunnel-Password
23	Framed-IPX-Network	70	ARAP-Password
24	State	71	ARAP-Features
25	Class	72	ARAP-Zone-Access
26	Vendor-Specific	73	ARAP-Security
27	Session-Timeout	74	ARAP-Security-Data
28	Idle-Timeout	75	Password-Retry
29	Termination-Action	76	Prompt
30	Called-Station-Id	77	Connect-Info
31	Calling-Station-Id	78	Configuration-Token
32	NAS-Identifier	79	EAP-Message
33	Proxy-State	80	Message-Authenticator
34	Login-LAT-Service	81	Tunnel-Private-Group-id
35	Login-LAT-Node	82	Tunnel-Assignment-id
36	Login-LAT-Group	83	Tunnel-Preference
37	Framed-AppleTalk-Link	84	ARAP-Challenge-Response
38	Framed-AppleTalk-Network	85	Acct-Interim-Interval
39	Framed-AppleTalk-Zone	86	Acct-Tunnel-Packets-Lost
40	Acct-Status-Type	87	NAS-Port-Id
41	Acct-Delay-Time	88	Framed-Pool
42	Acct-Input-Octets	89	(unassigned)
43	Acct-Output-Octets	90	Tunnel-Client-Auth-id
44	Acct-Session-Id	91	Tunnel-Server-Auth-id



The attribute types listed in Table 59 are defined by RFC 2865, RFC 2866, RFC 2867, and RFC 2568.

RADIUS extended attributes

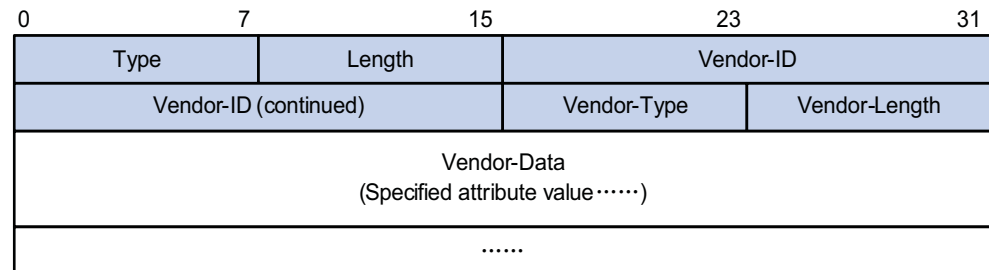
The RADIUS protocol features excellent extensibility. Attribute 26 (Vendor-Specific) defined by RFC 2865 allows a vendor to define extended attributes to implement functions that the standard RADIUS protocol does not provide.

A vendor can encapsulate multiple type-length-value (TLV) sub-attributes in RADIUS packets for extension in applications. As shown in Figure 231, a sub-attribute that can be encapsulated in Attribute 26 consists of the following four parts:

- Vendor-ID (four bytes): Indicates the ID of the vendor. Its most significant byte is 0 and the other three bytes contain a code complying with RFC 1700. The vendor ID of 3Com is 2011.

- Vendor-Type: Indicates the type of the sub-attribute.
- Vendor-Length: Indicates the length of the sub-attribute.
- Vendor-Data: Indicates the contents of the sub-attribute.

Figure 231 Segment of a RADIUS packet containing an extended attribute



Introduction to HWTACACS

3Com Terminal Access Controller Access Control System (HWTACACS) is an enhanced security protocol based on TACACS (RFC 1492). Similar to RADIUS, it uses the server/client model for information exchange between NAS and HWTACACS server.

HWTACACS implements AAA mainly for such users as Point-to-Point Protocol (PPP) users, Virtual Private Dial-up Network (VPDN) users, and terminal users. In a typical HWTACACS application, a terminal user needs to log onto the device for operations. Working as the HWTACACS client, the device sends the username and password to the HWTACACS sever for authentication. After passing authentication and being authorized, the user can log into the device to perform operations.

Differences between HWTACACS and RADIUS

HWTACACS and RADIUS have many common features, like implementing AAA, using a client/server model, using shared keys for user information security and having good flexibility and extensibility. Meanwhile, they also have differences, as listed in Table 60.

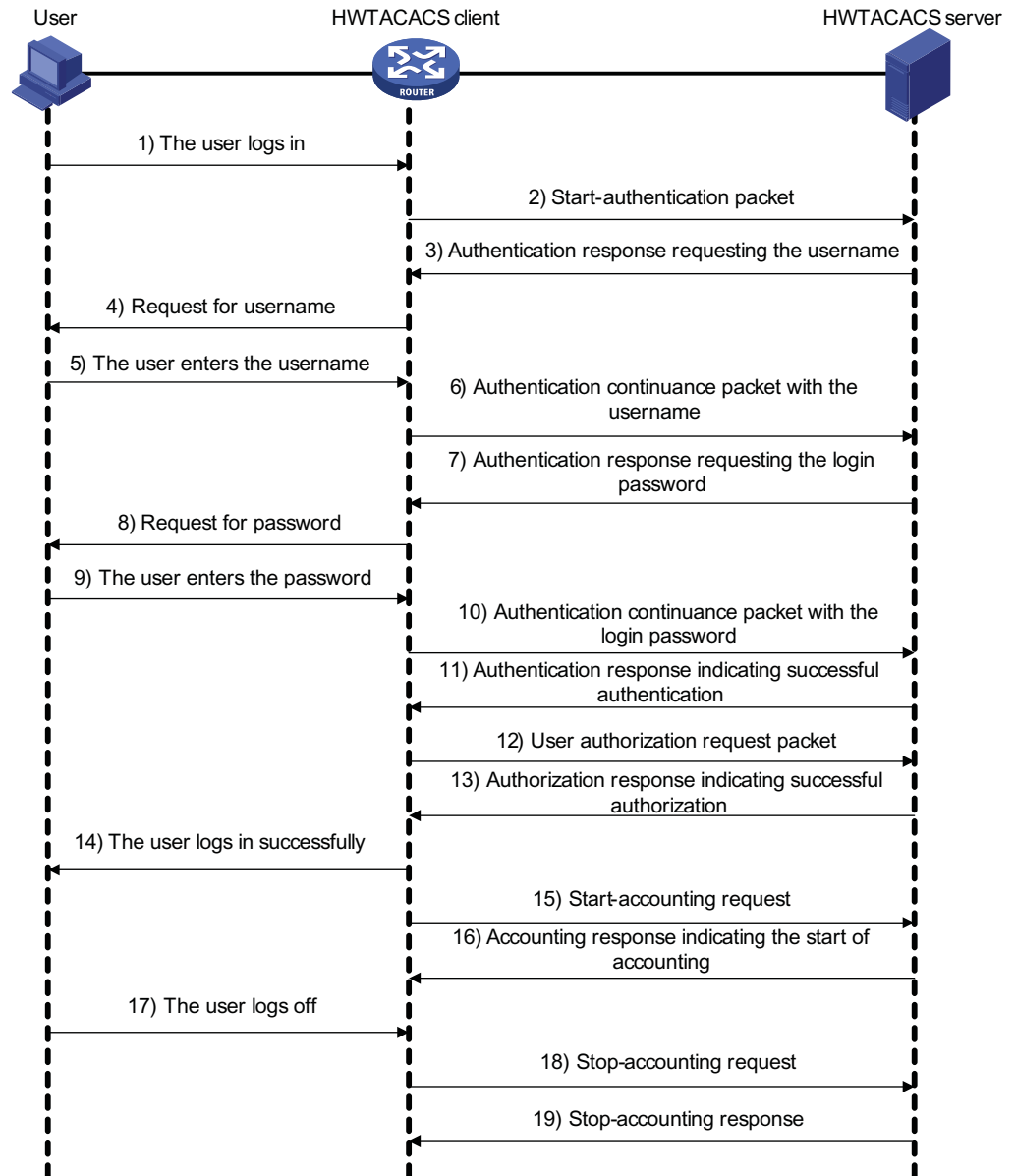
Table 60 Primary differences between HWTACACS and RADIUS

HWTACACS	RADIUS
Uses TCP, providing more reliable network transmission	Uses UDP, providing higher transport efficiency
Encrypts the entire packet except for the HWTACACS header	Encrypts only the password field in an authentication packet
Protocol packets are complicated and authorization is independent of authentication. Authentication and authorization can be deployed on different HWTACACS servers.	Protocol packets are simple and authorization is combined with authentication.
Supports authorized use of configuration commands. For example, an authenticated login user can be authorized to configure the device.	Does not support authorized use of configuration commands.

Basic message exchange process of HWTACACS

The following takes Telnet user as an example to describe how HWTACACS performs user authentication, authorization, and accounting. Figure 232 illustrates the basic message exchange process of HWTACACS.

Figure 232 Basic message exchange process of HWTACACS for a Telnet user



- 1 A Telnet user applies to access the NAS.
- 2 Upon receiving the request, the HWTACACS client sends a start-authentication packet to the HWTACACS server.
- 3 The HWTACACS server sends back an authentication response requesting the username.
- 4 Upon receiving the request, the HWTACACS client asks the user for the username.
- 5 The user enters the username.

- 6 After receiving the username from the user, the HWTACACS client sends to the server a continue-authentication packet carrying the username.
- 7 The HWTACACS server sends back an authentication response, requesting the login password.
- 8 Upon receipt of the response, the HWTACACS client requests of the user the login password.
- 9 The user enters the password.
- 10 After receiving the login password, the HWTACACS client sends to the HWTACACS server a continue-authentication packet carrying the login password.
- 11 The HWTACACS server sends back an authentication response indicating that the user has passed authentication.
- 12 The HWTACACS client sends the user authorization request packet to the HWTACACS server.
- 13 The HWTACACS server sends back the authorization response, indicating that the user is authorized now.
- 14 Knowing that the user is now authorized, the HWTACACS client pushes the configuration interface of the NAS to the user.
- 15 The HWTACACS client sends a start-accounting request to the HWTACACS server.
- 16 The HWTACACS server sends back an accounting response, indicating that it has received the start-accounting request.
- 17 The user logs off.
- 18 The HWTACACS client sends a stop-accounting request to the HWTACACS server.
- 19 The HWTACACS server sends back a stop-accounting packet, indicating that the stop-accounting request has been received.

- Protocols and Standards** The protocols and standards related to AAA, RADIUS, and HWTACACS include:
- RFC 2865: Remote Authentication Dial In User Service (RADIUS)
 - RFC 2866: RADIUS Accounting
 - RFC 2867: RADIUS Accounting Modifications for Tunnel Protocol Support
 - RFC 2868: RADIUS Attributes for Tunnel Protocol Support
 - RFC 2869: RADIUS Extensions
 - RFC 1492: An Access Control Protocol, Sometimes Called TACACS

AAA/RADIUS/HWTACACS Configuration Task List

AAA configuration task list

Task	Remarks
"Creating an ISP Domain" on page 758	Required
"Configuring ISP Domain Attributes" on page 758	Optional

Task	Remarks
"Configuring an AAA Authentication Scheme for an ISP Domain" on page 759	Required For local authentication, refer to "Configuring Local User Attributes" on page 763. For RADIUS authentication, refer to "Configuring RADIUS" on page 765. For HWTACACS authentication, refer to "Configuring HWTACACS" on page 771.
"Configuring an AAA Authorization Scheme for an ISP Domain" on page 760	Optional
"Configuring an AAA Accounting Scheme for an ISP Domain" on page 762	Optional
"Configuring Local User Attributes" on page 763	Optional
"Tearing down User Connections Forcibly" on page 765	Optional

RADIUS configuration task list

Task	Remarks
"Creating a RADIUS Scheme" on page 765	Required
"Specifying the RADIUS Authentication/Authorization Servers" on page 765	Required
"Configuring the RADIUS Accounting Servers and Relevant Parameters" on page 766	Optional
"Setting the Shared Key for RADIUS Packets" on page 767	Required
"Setting the Maximum Number of RADIUS Request Retransmission Attempts" on page 767	Optional
"Setting the Supported RADIUS Server Type" on page 768	Optional
"Setting the Status of RADIUS Servers" on page 768	Optional
"Configuring Attributes Related to the Data Sent to the RADIUS Server" on page 769	Optional
"Setting Timers Regarding RADIUS Servers" on page 770	Optional
"Configuring RADIUS Accounting-on" on page 771	Optional
"Enabling the Listening Port of the RADIUS Client" on page 771	Optional

HWTACACS configuration task list

Task	Remarks
"Creating a HWTACACS scheme" on page 771	Required
"Specifying the HWTACACS Authentication Servers" on page 772	Required
"Specifying the HWTACACS Authorization Servers" on page 772	Optional
"Specifying the HWTACACS Accounting Servers" on page 773	Optional
"Setting the Shared Key for HWTACACS Packets" on page 773	Required
"Configuring Attributes Related to the Data Sent to the TACACS Server" on page 774	Optional
"Setting Timers Regarding HWTACACS Servers" on page 774	Optional

Configuring AAA

By configuring AAA, you can provide network access service for legal users, protect the networking devices, and avoid unauthorized access and bilking. In addition, you can configure ISP domains to perform AAA on accessing users.

In AAA, users are divided into lan-access users (such as 802.1x users and MAC authentication users), login users (such as SSH, Telnet, FTP, and terminal access users), and command line users (that is, command line authentication users). Except for command line users, you can configure separate authentication/authorization/accounting policies for all the other type of users. Command line users can be configured with authorization policy independently.

Configuration Prerequisites

For remote authentication, authorization, or accounting, you must create the RADIUS or HWTACACS scheme first.

- RADIUS scheme: Reference a configured RADIUS scheme to implement authentication/authorization and accounting. For RADIUS scheme configuration, refer to “Configuring RADIUS” on page 765.
- HWTACACS scheme: Reference a configured HWTACACS scheme to implement authentication/authorization and accounting. For HWTACACS scheme configuration, refer to “Configuring HWTACACS” on page 771.

Creating an ISP Domain

For the NAS, each accessing user belongs to an ISP domain. Up to 16 ISP domains can be configured on a NAS. If a user does not provide the ISP domain name, the system considers that the user belongs to the default ISP domain.

Follow these steps to create an ISP domain:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create an ISP domain and enter ISP domain view	domain <i>isp-name</i>	Required
Return to system view	quit	-
Specify the default ISP domain	domain default { disable enable <i>isp-name</i> }	Optional The system-default ISP domain named system by default



- *You cannot delete the default ISP domain unless you change it to a non-default ISP domain (with the **domain default disable** command) first.*
- *If a user enters a username without an ISP domain name, the device uses the authentication scheme for the default ISP domain to authenticate the user.*

Configuring ISP Domain Attributes

Follow these steps to configure ISP domain attributes:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create an ISP domain and enter ISP domain view	domain <i>isp-name</i>	Required

To do...	Use the command...	Remarks
Place the ISP domain to the state of active or blocked	state { active block }	Optional When created, an ISP is in the state of active by default, and users in the domain can request network services.
Specify the maximum number of users in the ISP domain	access-limit { disable enable <i>max-user-number</i> }	Optional No limit by default
Configure the idle cut function	idle-cut { disable enable <i>minute</i> }	Optional Disabled by default
Enable the self-service server localization function and specify the URL of the self-service server for changing user password	self-service-url { disable enable <i>url-string</i> }	Optional Disabled by default



A self-service RADIUS server, for example, CAMS, is required for the self-service server localization function. With the self-service function, a user can manage and control his or her accounting information or module number. A server with self-service software is a self-service server.

Configuring an AAA Authentication Scheme for an ISP Domain

In AAA, authentication, authorization, and accounting are three separate processes. Authentication refers to the interactive authentication process of username/password/user information during access or service request. The authentication process neither sends authorization information to a supplicant nor triggers any accounting. You can configure AAA to use only authentication. If you do not perform any authentication configuration, the system-default ISP domain uses the local authentication scheme.

Before configuring an authentication scheme, complete these three tasks:

- For RADIUS or HWTACACS authentication, configure the RADIUS or HWTACACS scheme to be referenced first. The local and none authentication modes do not require any scheme.
- Determine the access mode or service type to be configured. With AAA, you can configure an authentication scheme specifically for each access mode and service type, limiting the authentication protocols that can be used for access.
- Determine whether to configure an authentication scheme for all access modes or service types.

Follow these steps to configure an AAA authentication scheme for an ISP domain:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create an ISP domain and enter ISP domain view	domain <i>isp-name</i>	Required

To do...	Use the command...	Remarks
Specify the default authentication scheme for all types of users	authentication default { hwtacacs-scheme <i>hwtacacs-scheme-name</i> [local] local none radius-scheme <i>radius-scheme-name</i> [local] }	Optional local by default
Specify the authentication scheme for LAN access users	authentication lan-access { local none radius-scheme <i>radius-scheme-name</i> [local] }	Optional The default authentication scheme is used by default.
Specify the authentication scheme for login users	authentication login { hwtacacs-scheme <i>hwtacacs-scheme-name</i> [local] local none radius-scheme <i>radius-scheme-name</i> [local] }	Optional The default authentication scheme is used by default.



- The authentication scheme specified with the **authentication default** command is for all types of users and has a priority lower than that for a specific access mode.
- With a RADIUS authentication scheme configured, AAA accepts only the authentication result from the RADIUS server. The response from the RADIUS server does include the authorization information when the authentication is successful, but the authentication process ignores the information.
- With the **radius-scheme** *radius-scheme-name* **local** or **hwtacacs-scheme** *hwtacacs-scheme-name* **local** keyword and argument combination configured, the local scheme is the backup scheme and is used only when the RADIUS server or TACACS server is not available.
- If the primary authentication scheme is **local** or **none**, the system performs local authentication or does not perform any authentication, rather than uses the RADIUS or HWTACACS scheme.

Configuring an AAA Authorization Scheme for an ISP Domain

In AAA, authorization is a separate process at the same level as authentication and accounting. Its responsibility is to send authorization requests to the specified authorization server and to send authorization information to users authorized. Authorization scheme configuration is optional in AAA configuration.

If you do not perform any authorization configuration, the system-default domain uses the local authorization scheme. With the authorization scheme of **none**, the users are not required to be authorized, in which case an authenticated user has the default right. The default right is visiting (the lowest one) for EXEC users (that is, console users who use the console, AUX, or Telnet or SSH to connect to the device, such as Telnet or SSH users. Each connection of these types is called an EXEC user). The default right for FTP users is to use the root directory of the device.

Before configuring an authorization scheme, complete these three tasks:

- 1 For HWTACACS authorization, configure the HWTACACS scheme to be referenced first. For RADIUS authorization, the RADIUS authorization scheme must be same as the RADIUS authentication scheme; otherwise, it does not take effect.
- 2 Determine the access mode or service type to be configured. With AAA, you can configure an authorization scheme specifically for each access mode and service type, limiting the authorization protocols that can be used for access.
- 3 Determine whether to configure an authorization scheme for all access modes or service types.

Follow these steps to configure an AAA authorization scheme for an ISP domain:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create an ISP domain and enter ISP domain view	domain <i>isp-name</i>	Required
Specify the default authorization scheme for all types of users	authorization default { hwtacacs-scheme <i>hwtacacs-scheme-name</i> [local] local none radius-scheme <i>radius-scheme-name</i> [local] }	Optional local by default
Specify the authorization scheme for command line users	authorization command hwtacacs-scheme <i>hwtacacs-scheme-name</i>	Optional The default authorization scheme is used by default.
Specify the authorization scheme for LAN access users	authorization lan-access { local none radius-scheme <i>radius-scheme-name</i> [local] }	Optional The default authorization scheme is used by default.
Specify the authorization scheme for login users	authorization login { hwtacacs-scheme <i>hwtacacs-scheme-name</i> [local] local none radius-scheme <i>radius-scheme-name</i> [local] }	Optional The default authorization scheme is used by default.



- The authorization scheme specified with the **authorization default** command is for all types of users and has a priority lower than that for a specific access mode.
- RADIUS authorization is special in that it takes effect only when the RADIUS authorization scheme is the same as the RADIUS authentication scheme. In addition, if a RADIUS authorization fails, the error message returned to the NAS says that the server is not responding.
- With the **radius-scheme** *radius-scheme-name* **local** or **hwtacacs-scheme** *hwtacacs-scheme-name* **local** keyword and argument combination configured, the local scheme is the backup scheme and is used only when the RADIUS server or TACACS server is not available.

- If the primary authentication scheme is **local** or **none**, the system performs local authorization or does not perform any authorization, rather than uses the RADIUS or HWTACACS scheme.
- Authorization information of the RADIUS server is sent to the RADIUS client along with the authorization response message; therefore, you cannot specify a separate RADIUS server. If you use RADIUS for authorization and authentication, you must use the same scheme setting for authorization and authentication; otherwise, the system will prompt you with an error message.

Configuring an AAA Accounting Scheme for an ISP Domain

In AAA, accounting is a separate process at the same level as authentication and authorization. Its responsibility is to send accounting start/update/end requests to the specified accounting server. Accounting is not required, and therefore accounting scheme configuration is optional. If you do not perform any accounting configuration, the system-default domain uses the local accounting scheme.

Before configuring an authorization scheme, complete these three tasks:

- 1 For RADIUS or HWTACACS accounting, configure the RADIUS or HWTACACS scheme to be referenced first. The local and none authentication modes do not require any scheme.
- 2 Determine the access mode or service type to be configured. With AAA, you can configure an accounting scheme specifically for each access mode and service type, limiting the accounting protocols that can be used for access.
- 3 Determine whether to configure an accounting scheme for all access modes or service types.

Follow these steps to configure an AAA accounting scheme for an ISP domain:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create an ISP domain and enter ISP domain view	domain <i>isp-name</i>	Required
Enable the accounting optional feature	accounting optional	Optional Disabled by default
Specify the default accounting scheme for all types of users	accounting default { hwtacacs-scheme <i>hwtacacs-scheme-name</i> [local] local none radius-scheme <i>radius-scheme-name</i> [local] }	Optional Local by default
Specify the accounting scheme for LAN access users	accounting lan-access { local none radius-scheme <i>radius-scheme-name</i> [local] }	Optional The default accounting scheme is used by default.

To do...	Use the command...	Remarks
Specify the accounting scheme for login users	accounting login { hwtacacs-scheme <i>hwtacacs-scheme-name</i> [local] local none radius-scheme <i>radius-scheme-name</i> [local] }	Optional The default accounting scheme is used by default.



- With the **accounting optional** command configured, a user that will be disconnected otherwise can use the network resources even when there is no available accounting server or the communication with the current accounting server fails.
- The accounting scheme specified with the **accounting default** command is for all types of users and has a priority lower than that for a specific access mode.
- With the **radius-scheme** *radius-scheme-name* **local** or **hwtacacs-scheme** *hwtacacs-scheme-name* **local** keyword and argument combination configured, the local scheme is the backup scheme and is used only when the RADIUS server or HWTACACS server is not available.
- If the primary accounting scheme is **local** or **none**, the system performs local accounting or does not perform any accounting, rather than uses the RADIUS or HWTACACS scheme.
- With the access mode of login, accounting is not supported for FTP services.

Configuring Local User Attributes

For local authentication, you must create a local user and configure the attributes.

A local user represents a set of users configured on a device, which are uniquely identified by the username. For a user requesting network service to pass local authentication, you must add an entry as required in the local user database of the device.

Follow these steps to configure the attributes for a local user:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Set the password display mode for all local users	local-user password-display-mode { auto cipher-force }	Optional auto by default
Add a local user and enter local user view	local-user <i>user-name</i>	Required No local user is configured by default
Configure a password for the local user	password { cipher simple } <i>password</i>	Required
Place the local user to the state of active or blocked	state { active block }	Optional When created, a local user is in the state of active by default, and the user can request network services.

To do...		Use the command...	Remarks
Specify the service types for the user	Specify the service types for the user	service-type { lan-access ssh telnet terminal } * [level <i>level</i>] }	Required No service is authorized to a user by default
	Authorize the user to use the FTP service and specify a directory for the user to access	service-type ftp [ftp-directory <i>directory</i>]	Optional By default, no service is authorized to a user and anonymous access to FTP service is not allowed. If you authorize a user to use the FTP service but do not specify a directory that the user can access, the user can access the root directory of the device by default.
Set the directory accessible to FTP/SFTP users		work-directory <i>directory-name</i>	Optional By default, FTP/SFTP users can access the root directory.
Set the priority level of the user		level <i>level</i>	Optional 0 by default
Set attributes for a LAN access user		attribute { access-limit <i>max-user-number</i> idle-cut <i>minute</i> ip <i>ip-address</i> location { nas-ip <i>ip-address</i> port <i>slot-number</i> <i>subslot-number port-number</i> port <i>slot-number</i> <i>subslot-number</i> <i>port-number</i> } mac <i>mac-address</i> vlan <i>vlan-id</i> } *	Optional If the user is bound to a remote port, the nas-ip parameter must be specified. If the user is bound to a local port, the nas-ip parameter does not need to be specified. The default value of nas-ip is 127.0.0.1, meaning the current host.



- With the **local-user password-display-mode cipher-force** command configured, a local user password is always displayed in cipher text, regardless of the configuration of the **password** command. In this case, if you use the **save** command to save the configuration, all existing local user passwords will still be displayed in cipher text after the device restarts, even if you restore the display mode to **auto**.
- Local authentication checks the service types of a local user. If the service types are not available, the user cannot pass authentication. During authorization, a user with no service type configured is authorized with no service by default.
- If you specify an authentication method that requires the username and password, including local authentication, RADIUS authentication and HWTACACS authentication, the level of the commands that a user can use after logging in depends on the priority of the user, or the priority of user interface level as with other authentication methods. For an SSH user using RSA public key authentication, the commands that can be used depend on the level configured on the user interface. For details regarding authentication method and command level, refer to “Controlling Login Users” on page 75 and “Configuring User Levels and Command Levels” on page 1026 respectively.
- Both the **service-type** and **level** commands can be used to specify user priority. The one used later has the final effect.

- The **attribute ip** command only applies to authentications that support IP address passing, such as 802.1x. If you configure the command to authentications that do not support IP address passing, such as MAC address authentication, the local authentication will fail.
- The **attribute port** command binds a port by its number only, regardless of the port type.
- The **idle-cut** command configured under ISP view applies to lan-access users only.

Tearing down User Connections Forcibly

Follow these steps to tear down user connections forcibly:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Tear down AAA user connections forcibly	cut connection { access-type { dot1x mac-authentication portal } all domain <i>isp-name</i> interface <i>interface-type interface-number</i> ip <i>ip-address</i> mac <i>mac-address</i> ucibindex <i>ucib-index</i> user-name <i>user-name</i> vlan <i>vlan-id</i> } [slot <i>slot-number</i>]	Required Applies to only LAN access user connections at present.

Configuring RADIUS

The RADIUS protocol is configured scheme by scheme. After creating a RADIUS scheme, you need to configure the IP addresses and UDP ports of the RADIUS servers for the scheme. The servers include authentication/authorization servers and accounting servers, or from another point of view, primary servers and secondary servers. In another words, the attributes of a RADIUS scheme mainly include IP addresses of primary and secondary servers, shared key, and RADIUS server type.

Actually, the RADIUS protocol configurations only set the parameters necessary for the information interaction between a NAS and a RADIUS server. For these settings to take effect, you must reference the RADIUS scheme containing those settings in ISP domain view. For information about the commands for referencing a scheme, refer to "Configuring AAA" on page 758.

Creating a RADIUS Scheme

Before performing other RADIUS configurations, follow these steps to create a RADIUS scheme and enter RADIUS scheme view:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a RADIUS scheme and enter RADIUS scheme view	radius scheme <i>radius-scheme-name</i>	Optional Not defined by default



A RADIUS scheme can be referenced by more than one ISP domain at the same time.

Specifying the RADIUS Authentication/Authorization Servers

Follow these steps to specify the RADIUS authentication/authorization servers:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Create a RADIUS scheme and enter RADIUS scheme view	radius scheme <i>radius-scheme-name</i>	Required Not defined by default
Configure the IP address and UDP port of the primary RADIUS authentication/authorization server	primary authentication <i>ip-address [port-number]</i>	Required The defaults are as follows: 0.0.0.0 for the IP address, and 1812 for the port.
Configure the IP address and UDP port of the secondary RADIUS authentication/authorization server	secondary authentication <i>ip-address [port-number]</i>	Optional The defaults are as follows: 0.0.0.0 for the IP address, and 1812 for the port.



- *In practice, you may specify two RADIUS servers as the primary and secondary authentication/authorization servers respectively. At a moment, a server can be the primary authentication/authorization server for a scheme and the secondary authentication/authorization servers for another scheme.*
- *The IP addresses of the primary and secondary authentication/authorization servers for a scheme cannot be the same. Otherwise, the configuration fails.*

Configuring the RADIUS Accounting Servers and Relevant Parameters

Follow these steps to specify the RADIUS accounting servers and perform related configurations:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a RADIUS scheme and enter RADIUS scheme view	radius scheme <i>radius-scheme-name</i>	Required Not defined by default
Configure the IP address and UDP port of the primary RADIUS accounting server	primary accounting <i>ip-address [port-number]</i>	Required The defaults are as follows: 0.0.0.0 for the IP address, and 1813 for the port.
Configure the IP address and UDP port of the secondary RADIUS accounting server	secondary accounting <i>ip-address [port-number]</i>	Optional The defaults are as follows: 0.0.0.0 for the IP address, and 1813 for the port.
Enable the device to buffer stop-accounting requests getting no responses	stop-accounting-buffer enable	Optional Enabled by default
Set the maximum number of stop-accounting request transmission attempts	retry stop-accounting <i>retry-times</i>	Optional 500 by default
Set the maximum number of accounting request transmission attempts	retry realtime-accounting <i>retry-times</i>	Optional 5 by default



- *In practice, you can specify two RADIUS servers as the primary and secondary accounting servers respectively; or specify one server to function as both. Besides, because RADIUS uses different UDP ports to receive*

authentication/authorization and accounting packets, the port for authentication/authorization must be different from that for accounting.

- You can set the maximum number of stop-accounting request transmission buffer, allowing the device to buffer and resend a stop-accounting request until it receives a response or the number of transmission retries reaches the configured limit. In the latter case, the device discards the packet.
- You can set the maximum number of accounting request transmission attempts on the device, allowing the device to disconnect a user when the number of accounting request transmission attempts for the user reaches the limit but it still receives no response to the accounting request.
- The IP addresses of the primary and secondary accounting servers cannot be the same. Otherwise, the configuration fails.
- Currently, RADIUS does not support keeping accounts on FTP users.

Setting the Shared Key for RADIUS Packets

The RADIUS client and RADIUS server use the MD5 algorithm to encrypt packets exchanged between them and a shared key to verify the packets. Only when the same key is used can they properly receive the packets and make responses.

Follow these steps to set the shared key for RADIUS packets:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a RADIUS scheme and enter RADIUS scheme view	radius scheme <i>radius-scheme-name</i>	Required Not defined by default
Set the shared key for RADIUS authentication/authorization or accounting packets	key { accounting authentication } string	Required No key by default



The shared key configured on the device must be the same as that configured on the RADIUS server.

Setting the Maximum Number of RADIUS Request Retransmission Attempts

Because RADIUS uses UDP packets to carry data, the communication process is not reliable. If a NAS receives no response from the RADIUS server before the response timeout timer expires, it is required to retransmit the RADIUS request. If the number of transmission attempts exceeds the specified limit but it still receives no response, it considers the authentication a failure.

Follow these steps to set the maximum number of RADIUS request retransmission attempts:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a RADIUS scheme and enter RADIUS scheme view	radius scheme <i>radius-scheme-name</i>	Required Not defined by default
Set the number of retransmission attempts of RADIUS packets	retry <i>retry-times</i>	Optional 3 by default



- The maximum number of retransmission attempts of RADIUS packets multiplied by the RADIUS server response timeout period cannot be greater than 75.
- Refer to the **timer response-timeout** command in the command manual for configuring RADIUS server response timeout period.

Setting the Supported RADIUS Server Type

Follow these steps to set the supported RADIUS server type:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a RADIUS scheme and enter RADIUS scheme view	radius scheme <i>radius-scheme-name</i>	Required Not defined by default
Specify the RADIUS server type supported by the device	server-type { extended standard }	Optional By default, the RADIUS server type is standard .



- If you change the type of RADIUS server, the data stream destined to the original RADIUS server will be restored to the default unit.
- When a third-party RADIUS is used, you can configure the RADIUS server to **standard** or **extended**. When CAMS server is used, you must RADIUS server to **extended**.

Setting the Status of RADIUS Servers

When a primary server, authentication/authorization server or accounting server, fails, the device automatically turns to the secondary server.

When both the primary and secondary servers are available, the device sends request packets to the primary server.

Once the primary server fails, the primary server turns into the state of block, and the device turns to the secondary server. In this case:

- If the secondary server is available, the device triggers the primary server quiet timer. After the quiet timer times out, the status of the primary server is active again and the status of the secondary server remains the same.
- If the secondary server fails, the device restores the status of the primary server to active immediately.

If the primary server has resumed, the device turns to use the primary server and stops communicating with the secondary server. After accounting starts, the communication between the client and the secondary server remains unchanged.

Follow these steps to set the status of RADIUS servers:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a RADIUS scheme and enter RADIUS scheme view	radius scheme <i>radius-scheme-name</i>	Required Not defined by default

To do...	Use the command...	Remarks
Set the status of the primary RADIUS authentication/authorization server	state primary authentication { active block }	Optional active for every server configured with IP address in the RADIUS scheme
Set the status of the primary RADIUS accounting server	state primary accounting { active block }	
Set the status of the secondary RADIUS authentication/authorization server	state secondary authentication { active block }	
Set the status of the secondary RADIUS accounting server	state secondary accounting { active block }	



- *If both the primary server and the secondary server are in the blocked state, it is necessary to manually turn the secondary server to the active state so that the secondary server can perform authentication. If the secondary server is still in the blocked state, the primary/secondary switchover cannot take place.*
- *If one server is in the active state while the other is blocked, the primary/secondary switchover will not take place even if the active server is not reachable.*

Configuring Attributes Related to the Data Sent to the RADIUS Server

Follow these steps to configure the attributes related to the data sent to the RADIUS server:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the RADIUS trap function	radius trap { accounting-server-down authentication-server-down }	Optional Disabled by default
Create a RADIUS scheme and enter RADIUS scheme view	radius scheme <i>radius-scheme-name</i>	Required Not defined by default
Specify the format of the username to be sent to a RADIUS server	user-name-format { with-domain without-domain }	Optional By default, the ISP domain name is included in the username.
Specify the unit for data flows or packets to be sent to a RADIUS server	data-flow-format { data byte giga-byte kilo-byte mega-byte } packet { giga-packet kilo-packet mega-packet one-packet } }	Optional The defaults are as follows: byte for data flows, and one-packet for data packets.
Set the source IP address of the device to send RADIUS packets	In RADIUS scheme view nas-ip <i>ip-address</i> In system view quit radius nas-ip <i>ip-address</i>	Use either command By default, the outbound port serves as the source IP address to send RADIUS packets



- *Some earlier RADIUS servers cannot recognize usernames that contain an ISP domain name, therefore before sending a username including a domain name to such a RADIUS server, the device must remove the domain name. This*

command is thus provided for you to decide whether to include a domain name in a username to be sent to a RADIUS server.

- If a RADIUS scheme defines that the username is sent without the ISP domain name, do not apply the RADIUS scheme to more than one ISP domain, thus avoiding the confused situation where the RADIUS server regards two users in different ISP domains but with the same userid as one.
- The **nas-ip** command in RADIUS scheme view is only for the current RADIUS scheme, while the **radius nas-ip** command in system view is for all RADIUS schemes. However, the **nas-ip** command in RADIUS scheme view overwrites the configuration of the **radius nas-ip** command.

Setting Timers Regarding RADIUS Servers

There are three timers regarding RADIUS servers:

- RADIUS server response timeout (**response-timeout**): If a NAS receives no response from the RADIUS server in a period of time after sending a RADIUS request (authentication/authorization or accounting request), it has to resend the request so that the user has more opportunity to obtain the RADIUS service. The NAS uses the RADIUS server response timeout timer to control the transmission interval.
- Primary server quiet timer (**timer quiet**): If the primary server is not reachable, its state changes to blocked, and the device will communicate with the secondary server with an IP address configured. If the secondary server is reachable, the primary server will resume active after the period specified by this timer, and the secondary server's state does not change.
- Real-time accounting interval (**realtime-accounting**): This timer defines the interval for performing real-time accounting of users. After this timer is set, the switch will send accounting information of online users to the RADIUS server at the specified interval.

Follow these steps to set timers regarding RADIUS servers:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a RADIUS scheme and enter RADIUS scheme view	radius scheme <i>radius-scheme-name</i>	Required Not defined by default
Set the RADIUS server response timeout timer	timer response-timeout <i>seconds</i>	Optional 3 seconds by default
Set the quiet timer for the primary server	timer quiet <i>minutes</i>	Optional 5 minutes by default
Set the real-time accounting interval	timer realtime-accounting <i>minutes</i>	Optional 12 minutes by default



- *The product of the maximum number of retransmission attempts of RADIUS packets and the RADIUS server response timeout period cannot be greater than 75. This product is also the upper limit of the timeout time of different access modules.*
- *For an access module, the product of the RADIUS server response timeout period and the maximum number of retransmission attempts must be smaller than the timeout time.*

- To configure the maximum number of retransmission attempts of RADIUS packets, refer to the command **retry** in the command manual.

Configuring RADIUS Accounting-on

With the accounting-on function enabled, a device sends, whenever it reboots, accounting-on packets to the RADIUS server, requesting the server to force its users offline.

Once configured, the accounting-on function is executed as soon as the device restarts and completes its configuration. In case that the majority of the RADIUS servers (a device can be configured with 16 schemes at most) fail to respond to the accounting-on packets, the number of accounting-on packet retransmission attempts is too big, or the accounting-on packet retransmission interval is too long, the device will not handle AAA services until all these packets are retransmitted and all RADIUS servers have responded to accounting-on packets.

Follow these steps to configure accounting-on function of a RADIUS server:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create RADIUS scheme and enter its view	radius scheme <i>radius-scheme-name</i>	Required Not defined by default
Enable accounting-on	accounting-on enable	Required Disabled by default
Set the number of accounting-on packet retransmission attempts	accounting-on enable send <i>send-times</i>	Optional 5 times by default
Set the retransmission interval of accounting-on packets	accounting-on enable interval <i>seconds</i>	Optional 3 seconds by default



*If the system has no authentication scheme enabled with the accounting-on function when you execute the **accounting-on enable** command, you need to save the configuration and restart the device so that the command takes effect. Otherwise, the command takes effect immediately.*

Enabling the Listening Port of the RADIUS Client

Follow these steps to enable the listening port of the RADIUS client:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the listening port of the RADIUS client	radius client enable	Optional Enabled by default

Configuring HWTACACS

Creating a HWTACACS scheme

The HWTACACS protocol is configured on a per scheme basis. Before performing other HWTACACS configurations, follow these steps to create a HWTACACS scheme and enter HWTACACS scheme view:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a HWTACACS scheme and enter HWTACACS scheme view	hwtacacs scheme <i>hwtacacs-scheme-name</i>	Required Not defined by default



- *Up to 16 HWTACACS schemes can be configured.*
- *A scheme can be deleted only when it is not referenced.*

Specifying the HWTACACS Authentication Servers

Follow these steps to specify the HWTACACS authentication servers:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a HWTACACS scheme and enter HWTACACS scheme view	hwtacacs scheme <i>hwtacacs-scheme-name</i>	Required Not defined by default
Configure the IP address and port of the primary HWTACACS authentication server	primary authentication <i>ip-address [port-number]</i>	Required The defaults are as follows: 0.0.0.0 for the IP address, and 49 for the TCP port.
Configure the IP address and port of the secondary HWTACACS authentication server	secondary authentication <i>ip-address [port-number]</i>	Required The defaults are as follows: 0.0.0.0 for the IP address, and 49 for the TCP port.



- *The IP addresses of the primary and secondary authentication servers cannot be the same. Otherwise, the configuration fails.*
- *You can remove an authentication server only when no active TCP connection for sending authentication packets is using it.*

Specifying the HWTACACS Authorization Servers

Follow these steps to specify the HWTACACS authorization servers:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a HWTACACS scheme and enter HWTACACS scheme view	hwtacacs scheme <i>hwtacacs-scheme-name</i>	Required Not defined by default
Configure the IP address and port of the primary HWTACACS authorization server	primary authorization <i>ip-address [port-number]</i>	Required The defaults are as follows: 0.0.0.0 for the IP address, and 49 for the TCP port.
Configure the IP address and port of the secondary HWTACACS authorization server	secondary authorization <i>ip-address [port-number]</i>	Required The defaults are as follows: 0.0.0.0 for the IP address, and 49 for the TCP port.



- The IP addresses of the primary and secondary authorization servers cannot be the same. Otherwise, the configuration fails.
- You can remove an authorization server only when no active TCP connection for sending authorization packets is using it.

Specifying the HWTACACS Accounting Servers

Follow these steps to specify the HWTACACS accounting servers and perform related configurations:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a HWTACACS scheme and enter HWTACACS scheme view	hwtacacs scheme <i>hwtacacs-scheme-name</i>	Required Not defined by default
Configure the IP address and port of the primary HWTACACS accounting server	primary accounting <i>ip-address [port-number]</i>	Required The defaults are as follows: 0.0.0.0 for the IP address, and 49 for the TCP port.
Configure the IP address and port of the secondary HWTACACS accounting server	secondary accounting <i>ip-address [port-number]</i>	Required The defaults are as follows: 0.0.0.0 for the IP address, and 49 for the TCP port.
Enable the device to buffer stop-accounting requests getting no responses	stop-accounting-buffer enable	Optional Enabled by default
Set the maximum number of stop-accounting request transmission attempts	retry stop-accounting <i>retry-times</i>	Optional 100 by default



- The IP addresses of the primary and secondary accounting servers cannot be the same. Otherwise, the configuration fails.
- You can remove an accounting server only when no active TCP connection for sending accounting packets is using it.
- Currently, HWTACACS does not support keeping accounts on FTP users.

Setting the Shared Key for HWTACACS Packets

When using a HWTACACS server as an AAA server, you can set a key to secure the communications between the device and the HWTACACS server.

The HWTACACS client and HWTACACS server use the MD5 algorithm to encrypt packets exchanged between them and a shared key to verify the packets. Only when the same key is used can they properly receive the packets and make responses.

Follow these steps to set the shared key for HWTACACS packets:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a HWTACACS scheme and enter HWTACACS scheme view	hwtacacs scheme <i>hwtacacs-scheme-name</i>	Required Not defined by default

To do...	Use the command...	Remarks
Set the shared keys for HWTACACS authentication, authorization, and accounting packets	key { accounting authentication authorization } <i>string</i>	Required No shared key exists by default.

Configuring Attributes Related to the Data Sent to the TACACS Server

Follow these steps to configure the attributes related to the data sent to the HWTACACS server:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a HWTACACS scheme and enter HWTACACS scheme view	hwtacacs scheme <i>hwtacacs-scheme-name</i>	Required Not defined by default
Specify the format of the username to be sent to a HWTACACS server	user-name-format { with-domain without-domain }	Optional By default, the ISP domain name is included in the username.
Specify the unit for data flows or packets to be sent to a HWTACACS server	data-flow-format { data byte giga-byte kilo-byte mega-byte } packet { giga-packet kilo-packet mega-packet one-packet } }	Optional The defaults are as follows: byte for data flows, and one-packet for data packets.
Set the source IP address of the device to send HWTACACS packets	In HWTACACS scheme view nas-ip <i>ip-address</i> In system view quit hwtacacs nas-ip <i>ip-address</i>	Use either command By default, the outbound port serves as the source IP address to send HWTACACS packets



- If a HWTACACS server does not support a username with the domain name, you can configure the device to remove the domain name before sending the username to the server.
- The **nas-ip** command in HWTACACS scheme view is only for the current HWTACACS scheme, while the **hwtacacs nas-ip** command in system view is for all HWTACACS schemes. However, the **nas-ip** command in HWTACACS scheme view overwrites the configuration of the **hwtacacs nas-ip** command.

Setting Timers Regarding HWTACACS Servers

Follow these steps to set timers regarding TACACS servers:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a HWTACACS scheme and enter HWTACACS scheme view	hwtacacs scheme <i>hwtacacs-scheme-name</i>	Required Not defined by default
Set the TACACS server response timeout timer	timer response-timeout <i>seconds</i>	Optional 5 seconds by default
Set the quiet timer for the primary server	timer quiet <i>minutes</i>	Optional 5 minutes by default

To do...	Use the command...	Remarks
Set the real-time accounting interval	timer realtime-accounting <i>minutes</i>	Optional 12 minutes by default



- For real-time accounting, a NAS must transmit the accounting information of online users to the HWTACACS accounting server periodically. Note that if the device does not receive any response to the information, it does not disconnect the online users forcibly
- The real-time accounting interval must be a multiple of 3.
- The setting of the real-time accounting interval somewhat depends on the performance of the NAS and the HWTACACS server: a shorter interval requires higher performance.

Displaying and Maintaining AAA/RADIUS/HWTACACS

Displaying and Maintaining AAA

To do...	Use the command...	Remarks
Display the configuration information of a specified ISP domain or all ISP domains	display domain [<i>isp-name</i>]	Available in any view
Display information about specified or all user connections	display connection [access-type { dot1x mac-authentication portal } domain <i>isp-name</i> interface <i>interface-type interface-number</i> ip <i>ip-address</i> mac <i>mac-address</i> ucibindex <i>ucib-index</i> user-name <i>user-name</i> vlan <i>vlan-id</i>]	Available in any view
Display information about specified or all local users	display local-user [idle-cut { disable enable } service-type { ftp lan-access ssh telnet terminal } state { active block } user-name <i>user-name</i> vlan <i>vlan-id</i>]	Available in any view

Displaying and Maintaining RADIUS

To do...	Use the command...	Remarks
Display the configuration information of a specified RADIUS scheme or all RADIUS schemes	display radius scheme [<i>radius-scheme-name</i>]	Available in any view
Display statistics about RADIUS packets	display radius statistics	Available in any view
Display information about buffered stop-accounting requests that get no responses	display stop-accounting-buffer { radius-scheme <i>radius-server-name</i> session-id <i>session-id</i> time-range <i>start-time stop-time</i> user-name <i>user-name</i> }	Available in any view
Clear RADIUS statistics	reset radius statistics	Available in user view

To do...	Use the command...	Remarks
Clear buffered stop-accounting requests that get no responses	reset stop-accounting-buffer { radius-scheme <i>radius-server-name</i> session-id <i>session-id</i> time-range <i>start-time stop-time</i> user-name <i>user-name</i> }	Available in user view
Clear the statistics on the local server	reset local-server statistics	Available in user view

Displaying and Maintaining HWTACACS

To do...	Use the command...	Remarks
Display configuration information or statistics of the specified or all HWTACACS schemes	display hwtacacs [<i>hwtacacs-server-name</i> [statistics]]	Available in any view
Display information about buffered stop-accounting requests that get no responses	display stop-accounting-buffer hwtacacs-scheme <i>hwtacacs-scheme-name</i>	Available in any view
Clear HWTACACS statistics	reset hwtacacs statistics { accounting all authentication authorization }	Available in user view
Clear buffered stop-accounting requests that get no responses	reset stop-accounting-buffer hwtacacs-scheme <i>hwtacacs-scheme-name</i>	Available in user view

AAA/RADIUS/HWTACACS Configuration Examples

AAA for Telnet Users by a HWTACACS Server

Network requirements

As shown in Figure 233, configure the switch to use the HWTACACS server to provide authentication, authorization, and accounting services to login users.

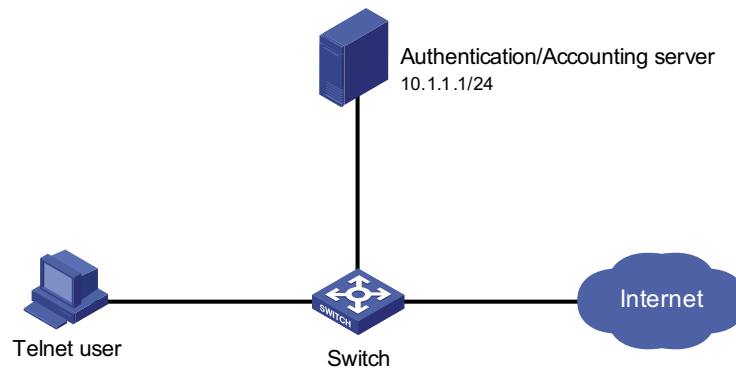
The HWTACACS server is used for authentication, authentication, and accounting. Its IP address is 10.1.1.1.

On the switch, set the shared keys for authentication, authorization, and accounting packets to **expert**. Configure the switch to remove the domain name from a user name before sending the user name to the HWTACACS server.

On the HWTACACS server, set the shared keys for packets exchanged with the switch to **expert**.

Network diagram

Figure 233 Configure AAA for Telnet users by a HWTACACS server



Configuration procedure

Configure the IP addresses of various interfaces (omitted).

Enable the Telnet server on the switch.

```
<Switch> system-view
[Switch] telnet server enable
```

Configure the switch to use AAA for Telnet users.

```
[Switch] user-interface vty 0 4
[Switch-ui-vty0-4] authentication-mode scheme
[Switch-ui-vty0-4] quit
```

Configure the HWTACACS scheme.

```
[Switch] hwtacacs scheme hwtac
[Switch-hwtacacs-hwtac] primary authentication 10.1.1.1 49
[Switch-hwtacacs-hwtac] primary authorization 10.1.1.1 49
[Switch-hwtacacs-hwtac] primary accounting 10.1.1.1 49
[Switch-hwtacacs-hwtac] key authentication expert
[Switch-hwtacacs-hwtac] key authorization expert
[Switch-hwtacacs-hwtac] key accounting expert
[Switch-hwtacacs-hwtac] user-name-format without-domain
[Switch-hwtacacs-hwtac] quit
```

Apply the AAA schemes to the domain.

```
[Switch] domain 1
[Switch-isp-1] authentication login hwtacacs-scheme hwtac
[Switch-isp-1] authorization login hwtacacs-scheme hwtac
[Switch-isp-1] accounting login hwtacacs-scheme hwtac
[Switch-isp-1] quit
```

You can achieve the same purpose by setting AAA schemes for all types of users.

```
[Switch] domain 1
[Switch-isp-1] authentication default hwtacacs-scheme hwtac
[Switch-isp-1] authorization default hwtacacs-scheme hwtac
```

```
[Switch-isp-1] accounting default hwtacacs-scheme hwtac
[Switch-isp-hwtacacs] accounting default hwtacacs-scheme hwtac
```

AAA for Telnet Users by Separate Servers

Network requirements

As shown in Figure 234, configure the switch to provide local authentication, HWTACACS authorization, and RADIUS accounting services to Telnet users. The user name and the password for Telnet users are both **telnet**.

The HWTACACS server is used for authorization. Its IP address is 10.1.1.2. On the switch, set the shared keys for packets exchanged with the TACACS server to **expert**. Configure the switch to remove the domain name from a user name before sending the user name to the HWTACACS server.

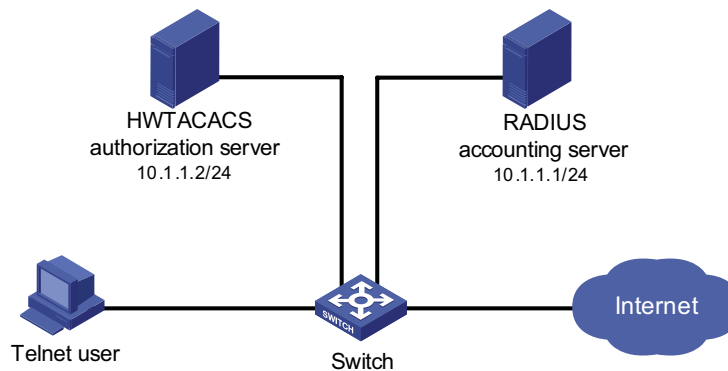
The RADIUS server is used for accounting. Its IP address is 10.1.1.1. On the switch, set the shared keys for packets exchanged with the RADIUS server to **expert**. Configure the switch to remove the domain name from a user name before sending the user name to the HWTACACS server.



Configuration of separate AAA for other types of users is similar to that given in this example. The only difference lies in the access type.

Network diagram

Figure 234 Configure AAA by separate servers for Telnet users



Configuration procedure

Configure the IP addresses of various interfaces (omitted).

Enable the Telnet server on the switch.

```
<Switch> system-view
[Switch] telnet server enable
```

Configure the switch to use AAA for Telnet users.

```
[Switch] user-interface vty 0 4
[Switch-ui-vty0-4] authentication-mode scheme
[Switch-ui-vty0-4] quit
```

Configure the HWTACACS scheme.

```
[Switch] hwtacacs scheme hwtac
[Switch-hwtacacs-hwtac] primary authorization 10.1.1.2 49
[Switch-hwtacacs-hwtac] key authorization expert
[Switch-hwtacacs-hwtac] user-name-format without-domain
[Switch-hwtacacs-hwtac] quit
```

Configure the RADIUS scheme.

```
[Switch] radius scheme rd
[Switch-radius-rd] primary accounting 10.1.1.1 1813
[Switch-radius-rd] key accounting expert
[Switch-radius-rd] server-type extended
[Switch-radius-rd] user-name-format without-domain
[Switch-radius-rd] quit
```

Create local user named telnet.

```
[Switch] local-user telnet
[Switch-luser-telnet] service-type telnet
[Switch-luser-telnet] password simple telnet
```

Configure the AAA schemes of the ISP domain.

```
[Switch] domain 1
[Switch-isp-1] authentication login local
[Switch-isp-1] authorization login hwtacacs-scheme hwtac
[Switch-isp-1] accounting login radius-scheme rd
[Switch-isp-1] quit
```

Configure the default AAA schemes for all types of users.

```
[Switch] domain 1
[Switch-isp-1] authentication default local
[Switch-isp-1] authorization default hwtacacs-scheme hwtac
[Switch-isp-1] accounting default radius-scheme cams
```

Troubleshooting AAA/RADIUS/HWTAC ACS

Troubleshooting RADIUS **Symptom1:** User authentication/authorization always fails.

Analysis:

- 1 A communication failure exists between the NAS and the RADIUS server.
- 2 The username is not in the format of *userid@isp-name* or no default ISP domain is specified for the NAS.
- 3 The user is not configured on the RADIUS server.
- 4 The password of the user is incorrect.
- 5 The RADIUS server and the NAS are configured with different shared key.

Solution:

Check that:

- 1 The NAS and the RADIUS server can ping each other.
- 2 The username is in the *userid@isp-name* format and a default ISP domain is specified on the NAS.
- 3 The user is configured on the RADIUS server.
- 4 The password entered by the user is correct.
- 5 The same shared key is configured on both the RADIUS server and the NAS.

Symptom2: RADIUS packets cannot reach the RADIUS server.

Analysis:

- 1 The communication link between the NAS and the RADIUS server is down (at the physical layer and data link layer).
- 2 The NAS is not configured with the IP address of the RADIUS server.
- 3 The UDP ports for authentication/authorization and accounting are not correct.

Solution:

Check that:

- 1 The communication links between the NAS and the RADIUS server work well at both physical and link layers.
- 2 The IP address of the RADIUS server is correctly configured on the NAS.
- 3 UDP ports for authentication/authorization/accounting configured on the NAS are the same as those configured on the RADIUS server.

Symptom3: A user is authenticated and authorized, but accounting for the user is not normal.

Analysis:

- 1 The accounting port number is not correct.
- 2 Configuration of the authentication/authorization server and the accounting server are not correct on the NAS. For example, one server is configured on the NAS to provide all the services of authentication/authorization and accounting, but in fact the services are provided by different servers.

Solution:

Check that:

- 1 The accounting port number is correctly set.
- 2 The authentication/authorization server and the accounting server are correctly configured on the NAS.

Troubleshooting HWTACACS

Refer to “Troubleshooting RADIUS” on page 779 if you encounter a HWTACACS fault.

54

ARP CONFIGURATION

When configuring ARP, go to these sections for information you are interested in:

- "ARP Overview" on page 781
- "Configuring ARP" on page 783
- "Configuring Gratuitous ARP" on page 785
- "Displaying and Maintaining ARP" on page 786

ARP Overview

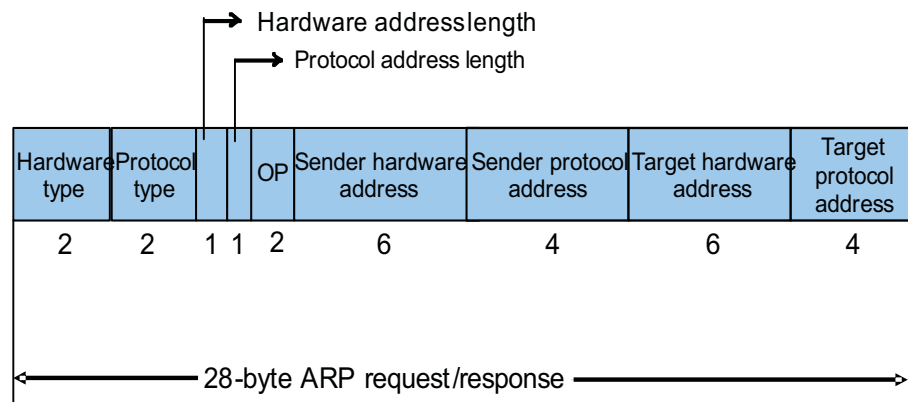
ARP Function Address Resolution Protocol (ARP) is used to resolve an IP address into a data link layer address.

An IP address is the address of a host at the network layer. To send a network layer packet to a destination host, the device must know the data link layer address (such as the MAC address) of the destination host. To this end, the IP address must be resolved into the corresponding data link layer address.



Unless otherwise stated, the data link layer addresses that appear in this chapter refer to the 48-bit Ethernet MAC addresses.

ARP Message Format **Figure 235** ARP message format



The following explains the fields in Figure 235.

- Hardware type: This field specifies the hardware address type. The value "1" represents Ethernet.
- Protocol type: This field specifies the type of the protocol address to be mapped. The hexadecimal value "0x0800" represents IP.

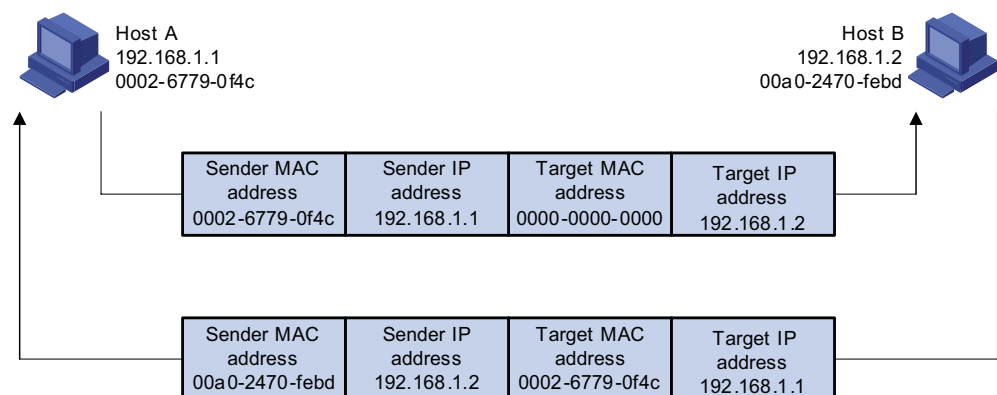
- Hardware address length and protocol address length: They respectively specify the length of a hardware address and a protocol address, in bytes. For an Ethernet address, the value of the hardware address length field is "6". For an IP(v4) address, the value of the protocol address length field is "4".
- OP: Operation code. This field specifies the type of ARP message. The value "1" represents an ARP request and "2" represents an ARP reply.
- Sender hardware address: This field specifies the hardware address of the device sending the message.
- Sender protocol address: This field specifies the protocol address of the device sending the message.
- Target hardware address: This field specifies the hardware address of the device the message is being sent to.
- Target protocol address: This field specifies the protocol address of the device the message is being sent to.

ARP Address Resolution Process

Suppose that Host A and Host B are on the same subnet and that Host A sends a message to Host B, as show in Figure 236. The resolution process is as follows:

- 1 Host A looks in its ARP mapping table to see whether there is an ARP entry for Host B. If Host A finds it, Host A uses the MAC address in the entry to encapsulate the IP packet into a data link layer frame and sends the frame to Host B.
- 2 If Host A finds no entry for Host B, Host A buffers the packet and broadcasts an ARP request, in which the source IP address and source MAC address are respectively the IP address and MAC address of Host A and the destination IP address and MAC address are respectively the IP address of Host B and an all-zero MAC address. Because the ARP request is sent in broadcast mode, all hosts on this subnet can receive the request, but only the requested host (namely, Host B) will process the request.
- 3 Host B compares its own IP address with the destination IP address in the ARP request. If they are the same, Host B saves the source IP address and source MAC address into its ARP mapping table, encapsulates its MAC address into an ARP reply, and unicasts the reply to Host A.
- 4 After receiving the ARP reply, Host A adds the MAC address of Host B into its ARP mapping table for subsequent packet forwarding. Meanwhile, Host A encapsulates the IP packet and sends it out.

Figure 236 ARP address resolution process



When Host A and Host B are not on the same subnet, Host A first sends an ARP request to the gateway. The destination IP address in the ARP request is the IP address of the gateway. After obtaining the MAC address of the gateway from an ARP reply, Host A encapsulates the packet and sends it to the gateway. Subsequently, the gateway broadcasts the ARP request, in which the destination IP address is the one of Host B. After obtaining the MAC address of Host B from another ARP reply, the gateway sends the packet to Host B.

ARP Mapping Table After obtaining the destination MAC address, the device adds the IP-to-MAC mapping into its own ARP mapping table. This mapping is used for forwarding packets with the same destination in future.

An ARP mapping table contains ARP entries, which fall into two categories: dynamic and static.

- 1 A dynamic entry is automatically created and maintained by ARP. It can get aged, be updated by a new ARP packet, or be overwritten by a static ARP entry. When the aging timer expires or the port goes down, the corresponding dynamic ARP entry will be removed.
- 2 A static ARP entry is manually configured and maintained. It cannot get aged or be overwritten by a dynamic ARP entry. It can be permanent or non-permanent.
 - A permanent static ARP entry can be directly used to forward packets. When configuring a permanent static ARP entry, you must configure a VLAN and outbound port for the entry besides the IP address and MAC address.
 - A non-permanent static ARP entry cannot be directly used for forwarding data. When configuring a non-permanent static ARP entry, you only need to configure the IP address and MAC address. When forwarding IP packets, the device sends an ARP request. If the source IP and MAC addresses in the received ARP reply are the same as the configured IP and MAC addresses, the device adds the port receiving the ARP reply into the static ARP entry. Now the entry can be used for forwarding IP packets.



Usually ARP dynamically implements and automatically seeks mappings from IP addresses to MAC addresses, without manual intervention.

Configuring ARP

Configuring a Static ARP Entry

A static ARP entry is effective when the device works normally. However, when a VLAN or VLAN interface to which a static ARP entry corresponds is deleted, the entry, if permanent, will be deleted, and if non-permanent and resolved, will become unresolved.

Follow these steps to configure a static ARP entry:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure a permanent static ARP entry	arp static <i>ip-address</i> <i>mac-address</i> <i>vlan-id</i> <i>interface-type</i> <i>interface-number</i>	Required No permanent static ARP entry is configured by default.

To do...	Use the command...	Remarks
Configure a non-permanent static ARP entry	arp static <i>ip-address mac-address</i>	Required No non-permanent static ARP entry is configured by default.



CAUTION: The *vlan-id* argument must be the ID of an existing VLAN which corresponds to the ARP entries. In addition, the Ethernet port following the argument must belong to that VLAN. A VLAN interface must be created for the VLAN.

Configuring the Maximum Number of ARP Entries for a VLAN Interface

Follow these steps to set the maximum number of dynamic ARP entries that a VLAN interface can learn:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter VLAN interface view	interface <i>Vlan-interface vlan-id</i>	-
Set the maximum number of dynamic ARP entries that a VLAN interface can learn	arp max-learning-num <i>number</i>	Optional 8192 by default.

Setting Aging Time for Dynamic ARP Entries

After dynamic ARP entries expire, the system will delete them from the ARP mapping table. You can adjust the aging time for dynamic ARP entries according to the actual network condition.

Follow these steps to set aging time for dynamic ARP entries:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Set aging time for dynamic ARP entries	arp timer aging <i>aging-time</i>	Optional 20 minutes by default.

Enabling the ARP Entry Check

The ARP entry check can control the device to learn multicast MAC addresses. With the ARP entry check enabled, the device cannot learn any ARP entry with a multicast MAC address. Configuring such a static ARP entry is not allowed either; otherwise, the system prompts error information.

After the ARP entry check is disabled, the device can learn the ARP entry with a multicast MAC address, and you can also configure such a static ARP entry on the device.

Follow these steps to enable the ARP entry check:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the ARP entry check	arp check enable	Optional Enabled by default.

ARP Configuration Example

Network requirements

- Enable the ARP entry check.
- Set the aging time for dynamic ARP entries to 10 minutes.
- Set the maximum number of dynamic ARP entries that VLAN-interface 10 can learn to 1000.
- Add a static ARP entry, with the IP address being 192.168.1.1/24, the MAC address being 000f-e201-0000, and the outbound port being GigabitEthernet 1/0/10 of VLAN 10.

Configuration procedure

```
<Sysname> system-view
[Sysname] arp check enable
[Sysname] arp timer aging 10
[Sysname] vlan 10
[Sysname-vlan10] port gigabitethernet 1/0/10
[Sysname-vlan10] quit
[Sysname] interface vlan-interface 10
[Sysname-vlan-interface10] arp max-learning-num 1000
[Sysname-vlan-interface10] quit
[Sysname] arp static 192.168.1.1 000f-e201-0000 10 gigabitethernet1/0/10
```

Configuring Gratuitous ARP

Introduction to Gratuitous ARP

A gratuitous ARP packet is a special ARP packet, in which the source IP address and destination IP address are both the IP address of the sender, the source MAC address is the MAC address of the sender, and the destination MAC address is a broadcast address.

A device can implement the following functions by sending gratuitous ARP packets:

- Determining whether its IP address is already used by another device.
- Informing other devices of its MAC address change so that they can update their ARP entries.

A device receiving a gratuitous ARP packet can add the information carried in the packet to its own dynamic ARP mapping table if it finds no corresponding ARP entry for the ARP packet in the cache.

Configuring Gratuitous ARP

Follow these steps to configure gratuitous ARP:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the device to send gratuitous ARP packets when receiving ARP requests from another network segment	gratuitous-arp-sending enable	Required By default, a device cannot send gratuitous ARP packets when receiving ARP requests from another network segment.

To do...	Use the command...	Remarks
Enable the gratuitous ARP packet learning function	gratuitous-arp-learning enable	Required Enabled by default.

Displaying and Maintaining ARP

To do...	Use the command...	Remarks
Display the ARP entries in the ARP mapping table	display arp { { all dynamic static } vlan <i>vlan-id</i> interface <i>interface-type interface-number</i> } [{ begin exclude include } <i>string</i> count]	Available in any view
Display the ARP entries for a specified IP address	display arp ip-address [{ begin exclude include } <i>string</i>]	Available in any view
Display the aging time for dynamic ARP entries	display arp timer aging	Available in any view
Clear ARP entries from the ARP mapping table	reset arp { all dynamic static interface <i>interface-type interface-number</i> }	Available in user view



Executing the **reset arp interface** *interface-type interface-number* command only removes dynamic ARP entries of the specified port. To remove specified static ARP entries, you need to use the **undo arp ip-address** command.

55

PROXY ARP CONFIGURATION

When configuring proxy ARP, go to these sections for information you are interested in:

- "Proxy ARP Overview" on page 787
- "Enabling Proxy ARP" on page 787
- "Displaying and Maintaining Proxy ARP" on page 787

Proxy ARP Overview

For an ARP request of a host on a network to be forwarded to an interface that is on the same network but isolated at Layer 2 or a host on another network, the device connecting the two physical or virtual networks must be able to respond to the request. This is achieved by proxy ARP.

Proxy ARP implements Layer 3 communication between VLAN interfaces isolated at Layer 2 or located on different networks.

In one of the following cases, you need to enable the local proxy ARP:

- Devices connected to different isolated Layer 2 ports in the same VLAN on a switch need to implement Layer 3 communication.
- With the `isolate-user-vlan` function enabled on a device attached to a switch, devices in different secondary VLANs need to implement Layer 3 communication.

Enabling Proxy ARP

Follow these steps to enable proxy ARP or enable local proxy ARP in VLAN interface view:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter VLAN interface view	interface Vlan-interface <i>vlan-id</i>	Required
Enable proxy ARP	proxy-arp enable	Required Disabled by default.
Enable local proxy ARP	local-proxy-arp enable	Required Disabled by default.

Displaying and Maintaining Proxy ARP

To do...	Use the command...	Remarks
Display whether proxy ARP is enabled	display proxy-arp [interface Vlan-interface <i>vlan-id</i>]	Available in any view

To do...	Use the command...	Remarks
Display whether local proxy ARP is enabled	display local-proxy-arp [interface Vlan-interface <i>vlan-id</i>]	Available in any view

Proxy ARP Configuration Examples

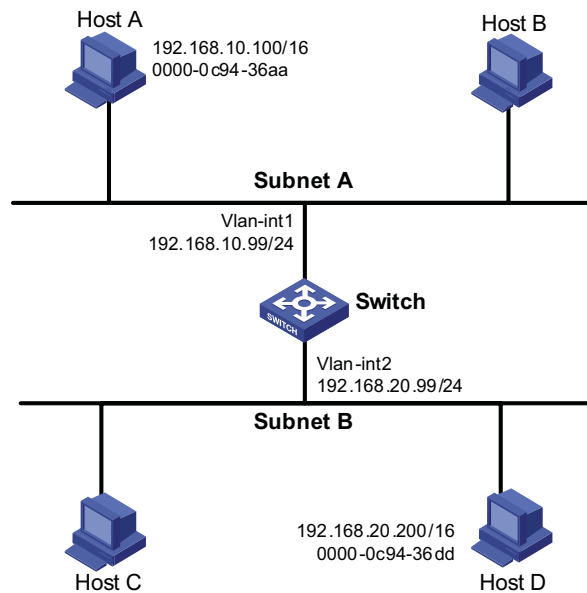
Proxy ARP Configuration Example

Network requirements

Host A and Host D have IP addresses of the same network segment. Host A belongs to VLAN 1, and Host D belongs to VLAN 2. Configure proxy ARP on the device to enable the communication between the two hosts.

Network diagram

Figure 237 Network diagram for proxy ARP



Configuration procedure

Configure Proxy ARP on the device to enable the communication between Host A and Host D.

```
<Switch> system-view
[Switch] vlan 2
[Switch-vlan2] quit
[Switch] interface vlan-interface 1
[Switch-Vlan-interface1] ip address 192.168.10.99 255.255.255.0
[Switch-Vlan-interface1] proxy-arp enable
[Switch-Vlan-interface1] quit
[Switch] interface vlan-interface 2
[Switch-Vlan-interface2] ip address 192.168.20.99 255.255.255.0
[Switch-Vlan-interface2] proxy-arp enable
[Switch-Vlan-interface2] quit
```

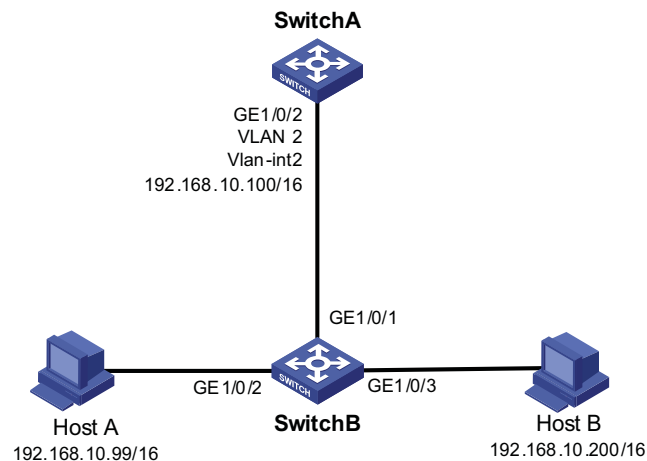

Local Proxy ARP Configuration Example in Case of Port Isolation

Network requirements

- Host A and Host B belong to the same VLAN, and are connected to GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3 of Switch B respectively.
- Switch B is connected to Switch A via GigabitEthernet 1/0/1.
- GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3 isolated at Layer 2 can implement Layer 3 communication.

Network diagram

Figure 238 Network diagram for local proxy ARP between isolated ports



Configuration procedure

1 Configure Switch B

Create VLAN 2 on Switch B, on which GigabitEthernet 1/0/1, GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3 belong to VLAN 2. Host A and Host B are isolated and unable to exchange Layer 2 packets.

```

<SwitchB> system-view
[SwitchB] vlan 2
[SwitchB-vlan2] port gigabitethernet 1/0/1
[SwitchB-vlan2] port gigabitethernet 1/0/2
[SwitchB-vlan2] port gigabitethernet 1/0/3
[SwitchB-vlan2] quit
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] port-isolate enable
[SwitchB-GigabitEthernet1/0/2] quit
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] port-isolate enable
[SwitchB-GigabitEthernet1/0/3] quit
  
```

2 Configure Switch A

Configure an IP address of VLAN-interface 2.

```

[SwitchA] vlan 2
[SwitchA-vlan2] port gigabitethernet 1/0/2
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 192.168.10.100 255.255.0.0
  
```

Ping Host B on Host A to verify that the two hosts cannot be pinged through, which indicates they are isolated at Layer 2.

Configure local proxy ARP to let Host A and Host B communicate at Layer 3.

```
[SwitchA-Vlan-interface2] local-proxy-arp enable  
[SwitchA-Vlan-interface2] quit
```

Ping Host B on Host A to verify that the two hosts can be pinged through, which indicates Layer 3 communication is implemented.

When configuring ARP, go to these sections for information you are interested in:

- “Introduction to DHCP” on page 791
- “DHCP Address Allocation” on page 792
- “DHCP Message Format” on page 793
- “DHCP Options” on page 794
- “Protocols and Standards” on page 796

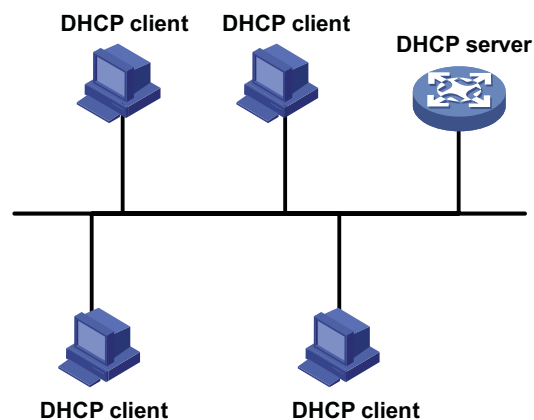
Introduction to DHCP

The fast expansion and growing complexity of networks result in scarce IP addresses assignable to hosts. Meanwhile, with the wide application of wireless networks, the frequent movement of laptops across networks requires that the IP addresses be changed accordingly. Therefore, related configurations on hosts become more complex. Dynamic Host Configuration Protocol (DHCP) was introduced to solve these problems.

DHCP is built on a client-server model, in which the client sends a configuration request and then the server returns a reply to send configuration parameters such as an IP address to the client.

A typical DHCP application, as shown in Figure 239, includes a DHCP server and multiple clients (PCs and laptops).

Figure 239 A typical DHCP application



When residing in a different subnet from the DHCP server, the DHCP client can get the IP address and other configuration parameters from the server via a DHCP relay agent. For information about the DHCP relay agent, refer to “Introduction to DHCP Relay Agent” on page 813.

DHCP Address Allocation

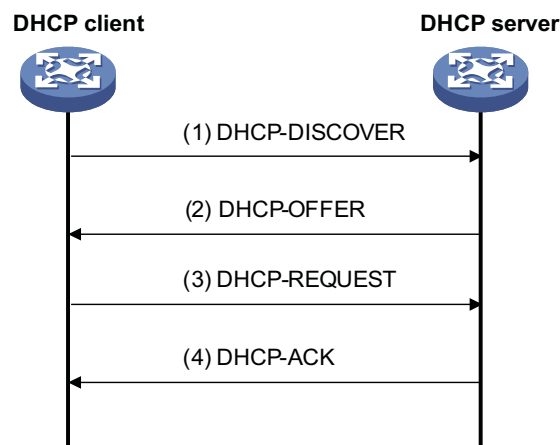
Allocation Mechanisms

DHCP supports three mechanisms for IP address allocation.

- Manual allocation: The network administrator assigns an IP address to a client like a WWW server, and DHCP conveys the assigned address to the client.
- Automatic allocation: DHCP assigns a permanent IP address to a client.
- Dynamic allocation: DHCP assigns an IP address to a client for a limited period of time, which is called a lease. Most clients obtain their addresses in this way.

Dynamic IP Address Allocation Process

Figure 240 Dynamic IP address allocation process



As shown in the figure above, a DHCP client obtains an IP address from a DHCP server via four steps:

- 1 The client broadcasts a DHCP-DISCOVER message to locate a DHCP server.
- 2 A DHCP server offers configuration parameters such as an IP address to the client in a DHCP-OFFER message. The sending mode of the DHCP-OFFER is determined by the flag field in the DHCP-DISCOVER message. Refer to “DHCP Message Format” on page 793 for related information.
- 3 If several DHCP servers send offers to the client, the client accepts the first received offer, and broadcasts it in a DHCP-REQUEST message to formally request the IP address.
- 4 All DHCP servers receive the DHCP-REQUEST message, but only the server to which the client sent a formal request for the offered IP address returns a DHCP-ACK message to the client, confirming that the IP address has been allocated to the client, or returns a DHCP-NAK unicast message, denying the IP address allocation.



- *After the client receives the DHCP-ACK message, it will probe whether the IP address assigned by the server is in use by broadcasting a gratuitous ARP packet. If the client receives no response within specified time, the client can use this IP address. Otherwise, the client sends a DHCP-DECLINE message to the server to request an IP address again.*

- If there are multiple DHCP servers, IP addresses offered by other DHCP servers are assignable to other clients.

IP Address Lease Extension

The IP address dynamically allocated by a DHCP server to a client has a lease. After the lease duration elapses, the IP address will be reclaimed by the DHCP server. If the client wants to use the IP address again, it has to extend the lease duration.

After the half lease duration elapses, the DHCP client will send the DHCP server a DHCP-REQUEST unicast message to extend the lease duration. Upon availability of the IP address, the DHCP server returns a DHCP-ACK unicast confirming that the client's lease duration has been extended, or a DHCP-NAK unicast denying the request.

If the client receives the DHCP-NAK message, it will broadcast another DHCP-REQUEST message for lease extension after 7/8 lease duration elapses. The DHCP server will handle the request as above mentioned.

DHCP Message Format

Figure 241 gives the DHCP message format, which is based on the BOOTP message format and involves eight types. These types of messages have the same format except that some fields have different values. The numbers in parentheses indicate the size of each field in bytes.

Figure 241 DHCP message format

0	7	15	23	31
op (1)	htype (1)		hlen (1)	hops (1)
xid (4)				
secs (2)		flags (2)		
ciaddr (4)				
yiaddr (4)				
siaddr (4)				
giaddr (4)				
chaddr (16)				
sname (64)				
file (128)				
options (variable)				

- op: Message type defined in option field. 1 = REQUEST, 2 = REPLY
- htype,hlen: Hardware address type and length of a DHCP client.
- hops: Number of relay agents a request message traveled.
- xid: Transaction ID, a random number chosen by the client to identify an IP address allocation.
- secs: Filled in by the client, the number of seconds elapsed since the client began address acquisition or renewal process. Currently this field is reserved and set to 0.
- flags: The leftmost bit is defined as the BROADCAST (B) flag. If this flag is set to 0, the DHCP server sent a reply back by unicast; if this flag is set to 1, the DHCP

server sent a reply back by broadcast. The remaining bits of the flags field are reserved for future use.

- ciaddr: Client IP address.
- yiaddr: 'your' (client) IP address, assigned by the server.
- siaddr: Server IP address, from which the clients obtained configuration parameters.
- giaddr: The first relay agent IP address a request message traveled.
- chaddr: Client hardware address.
- sname: The server host name, from which the client obtained configuration parameters.
- file: Bootfile name and routing information, defined by the server to the client.
- options: Optional parameters field that is variable in length, which includes the message type, lease, DNS IP address, WINS IP address and so forth.

DHCP Options

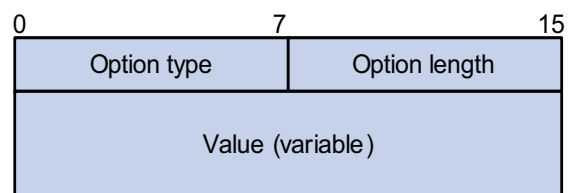
DHCP Options Overview

The DHCP message adopts the same format as the Bootstrap Protocol (BOOTP) message for compatibility, but differs from it in the option field, which identifies new features for DHCP.

DHCP uses the option field in DHCP messages to carry control information and network configuration parameters, implementing dynamic address allocation and providing more network configuration information for clients.

Figure 242 shows the DHCP option format.

Figure 242 DHCP option format



Introduction to DHCP Options

The common DHCP options are:

- Option 6: DNS server option. It specifies the DNS server IP address to be assigned to the client.
- Option 51: IP address lease option.
- Option 53: DHCP message type option. It identifies the type of the DHCP message.
- Option 55: Parameter request list option. It is used by a DHCP client to request specified configuration parameters. The option contains values that correspond to the parameters requested by the client.
- Option 66: TFTP server name option. It specifies a TFTP server to be assigned to the client.

- Option 67: Bootfile name option. It specifies the bootfile name to be assigned to the client.
- Option 150: TFTP server IP address option. It specifies the TFTP server IP address to be assigned to the client.

For more information about DHCP options, refer to RFC 2132.

Self-Defined Options

Some options have no unified definitions in RFC 2132. The formats of some self-defined options are introduced as follows.

Relay agent option (Option 82)

Option 82 is the relay agent option in the option field of the DHCP message. It records the location information of the DHCP client. When a DHCP relay agent receives a client's request, it adds Option 82 to the request message and sends it to the server.

The administrator can locate the DHCP client to further implement security control and accounting. The Option 82 supporting server can also use such information to define individual assignment policies of IP address and other parameters for the clients.

Option 82 involves at most 255 sub-options. At least one sub-option must be defined. Now the DHCP relay agent supports two sub-options: sub-option 1 (Circuit ID) and sub-option 2 (Remote ID).

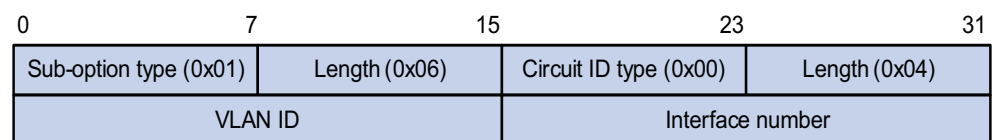
Option 82 has no unified definition. Its padding formats vary with vendors. Currently the device supports two padding formats: normal and verbose.

1 Normal padding format

The padding contents for sub-options in the normal padding format are:

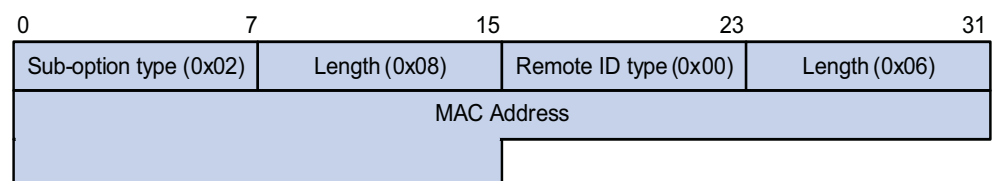
- sub-option 1: Padded with the VLAN ID and number of the port that received the client's request. The following figure gives its format. The value of the sub-option type is 1, and that of the circuit ID type is 0.

Figure 243 Sub-option 1 in normal padding format



- sub-option 2: Padded with the MAC address of the interface that received the client's request. The following figure gives its format. The value of the sub-option type is 2, and that of the remote ID type is 0.

Figure 244 Sub-option 2 in normal padding format



2 Verbose padding format:

The padding contents for sub-options in the verbose padding format are:

- sub-option 1: Padded with the user-specified access node identifier (ID of the device that adds Option 82 in DHCP messages), and type, number, and VLAN ID of the port that received the client's request. Its format is shown in the following figure.

Figure 245 Sub-option 1 in verbose padding format

Sub-option type (0x01)	Length	Node identifier
Port type		Port number
VLAN ID		



In the above figure, except that the VLAN ID field has a fixed length of 2 bytes, all the other padding contents of sub-option 1 are length variable.

- sub-option 2: Padded with the MAC address of the interface that received the client's request. It has the same format as that in normal padding format, as shown in Figure 244.

Option 184

Option 184 is a reserved option, and parameters in the option can be defined as needed. The device supports Option 184 carrying the voice related parameters, so a DHCP client with voice functions can get an IP address along with specified voice parameters from the DHCP server.

Option 184 involves the following sub-options:

- Sub-option 1: IP address of the primary network calling processor, which is a server serving as the network calling control source and providing program downloads.
- Sub-option 2: IP address of the backup network calling processor that DHCP clients will contact when the primary one is unreachable.
- Sub-option 3: Voice VLAN ID and the result whether DHCP clients take this ID as the voice VLAN or not.
- Sub-option 4: Failover route that specifies the destination IP address and the called number (SIP users use such IP addresses and numbers to communicate with each other) that a SIP user uses to reach another SIP user when both the primary and backup calling processors are unreachable.




You must define the sub-option 1 to make other sub-options take effect.

Protocols and Standards

- RFC2131: Dynamic Host Configuration Protocol
- RFC2132: DHCP Options and BOOTP Vendor Extensions
- RFC1542: Clarifications and Extensions for the Bootstrap Protocol
- RFC 3046: DHCP Relay Agent Information Option

DHCP SERVER CONFIGURATION

When configuring the DHCP server, go to these sections for information you are interested in:

- "Introduction to DHCP Server" on page 797
 - "DHCP Server Configuration Task List" on page 799
 - "Enabling DHCP" on page 799
 - "Enabling the DHCP Server on an Interface" on page 799
 - "Configuring an Address Pool for the DHCP Server" on page 800
 - "Configuring the DHCP Server Security Functions" on page 806
 - "Configuring the Handling Mode for Option 82" on page 808
 - "Displaying and Maintaining the DHCP Server" on page 808
 - "DHCP Server Configuration Examples" on page 809
 - "Troubleshooting DHCP Server Configuration" on page 811
-  ■ *The DHCP server configuration is supported only on VLAN interfaces and loopback interfaces. The secondary IP address pool configuration is not supported on loopback interfaces.*
- *DHCP Snooping must be disabled on the DHCP server.*

Introduction to DHCP Server

Application Environment

The DHCP server is well suited to the network where:

- It is hard to implement manual configuration and centralized management.
- The hosts are more than the assignable IP addresses and it is impossible to assign a fixed IP address to each host. For example, an ISP limits the number of hosts to access the Internet at a time, so lots of hosts need to acquire IP addresses dynamically.
- A few hosts need fixed IP addresses.

DHCP Address Pool **Address pool structure**

In response to a client's request, the DHCP server selects an idle IP address from an address pool and sends it together with other parameters such as lease and DNS server address to the client.

The address pool database is organized as a tree. The root of the tree is the address pool for natural networks, branches are address pools for subnets, and

leaves are addresses statically bound to clients. For the same level address pools, a previously configured pool has a higher selection priority than a new one.

At the very beginning, subnetworks inherit network parameters and clients inherit subnetwork parameters. Therefore, common parameters, for example a DNS server address, should be configured at the highest (network or subnetwork) level of the tree.

After establishment of the inheritance relationship, the new configuration at the higher level (father) of the tree will be:

- Inherited if the lower level (child) has no such configuration, or
- Overridden if the lower level (child) has such configuration.



The IP address lease does not enjoy the inheritance attribute.

Principles for selecting an address pool

The DHCP server observes the following principles to select an address pool to assign IP addresses to clients:

- 1 If there is an address pool where an IP address is statically bound to the MAC address or ID of the client, the DHCP server will select this address pool and assign the statically bound IP address to the client. For the configuration of this address pool, refer to section “Configuring manual address allocation” on page 800.
- 2 Otherwise, the DHCP server will select the smallest address pool that contains the IP address of the receiving interface (if the client and the server reside in the same network segment), or the smallest address pool that contains the IP address specified in the giaddr field of the client’s request (if a DHCP relay agent is in-between). If no IP address is available in such address pool, the DHCP server will fail to assign an address to the client because it cannot assign an IP address from the father address pool to the client. For the configuration of such address pool, refer to section “Configuring dynamic address allocation” on page 801.

For example, two address pools are configured on the DHCP server. The ranges of IP addresses that can be dynamically assigned are 1.1.1.0/24 and 1.1.1.0/25 respectively. If the IP address of the interface receiving DHCP requests is 1.1.1.1/25, the DHCP server will select IP addresses for clients from the 1.1.1.0/25 address pool. If no IP address is available in the 1.1.1.0/25 address pool, the DHCP server will fail to assign addresses to clients. If the IP address of the interface receiving DHCP requests is 1.1.1.130/25, the DHCP server will select IP addresses for clients from the 1.1.1.0/24 address pool.



Keep the IP addresses for dynamic allocation within the subnet where the interface of the DHCP server resides to avoid wrong IP address allocation.

IP Address Allocation Sequence

A DHCP server assigns an IP address to a client according to the following sequence:

- 1 The IP address manually bound to the client’s MAC address or ID
- 2 The IP address that was ever assigned to the client
- 3 The IP address designated by the Option 50 field in a DHCP-DISCOVER message
- 4 The first assignable IP address found in a proper DHCP address pool

- 5 The IP address that was a conflict or passed its lease duration
If no IP address is assignable, the server will not respond.

DHCP Server Configuration Task List

Complete the following tasks to configure the DHCP server:

Task	Remarks
"Enabling DHCP" on page 799	Required
"Enabling the DHCP Server on an Interface" on page 799	Optional
"Configuring an Address Pool for the DHCP Server" on page 800	Required
"Configuring the DHCP Server Security Functions" on page 806	Optional
"Configuring the Handling Mode for Option 82" on page 808	Optional

Enabling DHCP

Enable DHCP before performing other configurations.

Follow these steps to enable DHCP:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable DHCP	dhcp enable	Required Disabled by default.

Enabling the DHCP Server on an Interface

With the DHCP server enabled on an interface, upon receiving a client's request, the DHCP server will assign an IP address from its address pool to the DHCP client.

Follow these steps to enable the DHCP server on an interface:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable the DHCP server on an interface	dhcp select server global-pool [subaddress]	Optional Enabled by default.



The **subaddress** keyword is valid only when the server and client are on the same subnet. If a DHCP relay agent exists in between, regardless of **subaddress**, the DHCP server will select an IP address from the address pool of the subnet which contains the primary IP address of the DHCP relay agent's interface (connected to the client).

When the DHCP server and client are on the same subnet, the server will

- With **subaddress** specified, assign an IP address from the address pool of the subnet which the secondary IP address of the server's interface connected to the client belongs to, or assign from the first secondary IP address if several secondary IP addresses exist. If no secondary IP address is configured for the interface, the server is unable to assign an IP address to the client.

- Without **subaddress** specified, assign an IP address from the address pool of the subnet which the primary IP address of the server's interface (connected to the client) belongs to.

Configuring an Address Pool for the DHCP Server

Configuration Task List Complete the following tasks to configure an address pool:

Task	Remarks	
"Creating a DHCP Address Pool" on page 800	Required	
"Configuring an Address Allocation Mode" on page 800	"Configuring manual address allocation" on page 800 "Configuring dynamic address allocation" on page 801	Required to configure either of the two
"Configuring a Domain Name Suffix for the Client" on page 802		Optional
"Configuring DNS Servers for the Client" on page 802		
"Configuring WINS Servers and NetBIOS Node Type for the Client" on page 803		
"Configuring the BIMS Server Information for the Client" on page 803		
"Configuring Gateways for the Client" on page 804		
"Configuring Option 184 Parameters for the Client with Voice Service" on page 804		
"Configuring the TFTP Server and Bootfile Name for the Client" on page 805		
"Configuring Self-Defined DHCP Options" on page 805		

Creating a DHCP Address Pool

Follow these steps to create a DHCP address pool:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a DHCP address pool and enter its view	dhcp server ip-pool <i>pool-name</i>	Required No DHCP address pool is created by default.

Configuring an Address Allocation Mode



CAUTION: You can configure either the static binding or dynamic address allocation for an address pool as needed.

It is required to specify an address range for the dynamic address allocation. A static binding is a special address pool containing only one IP address.

Configuring manual address allocation

Some DHCP clients such as a WWW server need fixed IP addresses. You can create a static binding of a client's MAC or ID to IP address in the DHCP address pool.

When the client with the MAC address or ID requests an IP address, the DHCP server will find the IP address from the binding for the client.

A DHCP address pool now supports only one static binding, which can be a MAC-to-IP or ID-to-IP binding.

Follow these steps to configure the static binding in a DHCP address pool:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Enter DHCP address pool view		dhcp server ip-pool <i>pool-name</i>	-
Bind IP addresses statically		static-bind ip-address <i>ip-address</i> [<i>mask-length</i> mask <i>mask</i>]	Required No IP addresses are statically bound by default.
Bind MAC addresses or IDs statically	Specify the MAC address	static-bind mac-address <i>mac-address</i>	Required to configure either of the two
	Specify the ID	static-bind client-identifier <i>client-identifier</i>	Neither is bound statically by default.



- Use the **static-bind ip-address** command together with **static-bind mac-address** or **static-bind client-identifier** command to accomplish a static binding configuration.
- In a DHCP address pool, if you execute the **static-bind mac-address** command before the **static-bind client-identifier** command, the latter will overwrite the former and vice versa.
- If you use the **static-bind ip-address**, **static-bind mac-address**, or **static-bind client-identifier** command repeatedly in the DHCP address pool, the new configuration will overwrite the previous one.
- The IP address of the static binding cannot be an interface address of the DHCP server. Otherwise, an IP address conflict may occur and the bound client cannot obtain an IP address correctly.
- The ID of the static binding must be identical to the ID displayed by using the **display dhcp client verbose** command on the client. Otherwise, the client cannot obtain an IP address.

Configuring dynamic address allocation

You need to specify one and only one address range using a mask for the dynamic address allocation.

To avoid address conflicts, the DHCP server excludes IP addresses used by the GW, FTP server and so forth from dynamic allocation.

You can specify the lease duration for a DHCP address pool different from others, and a DHCP address pool can only have the same lease duration. A lease does not enjoy the inheritance attribute.

Follow these steps to configure the dynamic address allocation:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter DHCP address pool view	dhcp server ip-pool <i>pool-name</i>	-
Specify an IP address range	network <i>network-address</i> [<i>mask-length</i> mask <i>mask</i>]	Required Not specified by default, meaning no assignable address.
Specify the address lease duration	expired { day <i>day</i> [hour <i>hour</i> [minute <i>minute</i>]] unlimited }	Optional One day by default.
Return to system view	quit	-
Exclude IP addresses from automatic allocation	dhcp server forbidden-ip <i>low-ip-address</i> [<i>high-ip-address</i>]	Optional Except IP addresses of the DHCP server interfaces, all addresses in the DHCP address pool are assignable by default.



- In DHCP address pool view, using the **network** command repeatedly overwrites the previous configuration.
- Using the **dhcp server forbidden-ip** command repeatedly can specify multiple IP address ranges not assignable.

Configuring a Domain Name Suffix for the Client

You can specify a domain name suffix in each DHCP address pool on the DHCP server to provide the clients with the domain name suffix. With this suffix assigned, the client needs only input part of a domain name, and the system will add the domain name suffix for name resolution. For details about DNS, refer to “Configuring the DNS Client” on page 973.

Follow these steps to configure a domain name suffix in the DHCP address pool:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter the DHCP address pool view	dhcp server ip-pool <i>pool-name</i>	-
Specify a domain name suffix for the client	domain-name <i>domain-name</i>	Required Not specified by default.

Configuring DNS Servers for the Client

When a DHCP client wants to access a host on the Internet via the host name, it contacts a Domain Name System (DNS) server holding host name-to-IP address mappings to get the host IP address. You can specify up to eight DNS servers in the DHCP address pool.

Follow these steps to configure DNS servers in the DHCP address pool:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Enter DHCP address pool view	dhcp server ip-pool <i>pool-name</i>	-
Specify DNS servers for the client	dns-list <i>ip-address&<1-8></i>	Required Not specified by default.

Configuring WINS Servers and NetBIOS Node Type for the Client

A Microsoft DHCP client using NetBIOS protocol contacts a Windows Internet Naming Service (WINS) server for name resolution. Therefore, the DHCP server should assign a WINS server address when assigning an IP address to the client.

You can specify up to eight WINS servers in a DHCP address pool.

You need to specify in a DHCP address pool a NetBIOS node type for the client to approach name resolution. There are four NetBIOS node types:

- **b (broadcast)-node:** The b-node client sends the destination name in a broadcast message. The destination returns its IP address to the client after receiving the message.
- **p (peer-to-peer)-node:** The p-node client sends the destination name in a unicast message to the WINS server, and the WINS server returns the destination IP address.
- **m (mixed)-node:** A combination of broadcast first and peer-to-peer second. The m-node client broadcasts the destination name, if no response, then unicasts the destination name to the WINS server to get the destination IP address.
- **h (hybrid)-node:** A combination of peer-to-peer first and broadcast second. The h-node client unicasts the destination name to the WINS server, if no response, then broadcasts it to get the destination IP address.

Follow these steps to configure WINS servers and NetBIOS node type in the DHCP address pool:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter DHCP address pool view	dhcp server ip-pool <i>pool-name</i>	-
Specify WINS server IP addresses for the client	nbns-list <i>ip-address&<1-8></i>	Required (optional for b-node) No address is specified by default.
Specify the NetBIOS node type	netbios-type { b-node h-node m-node p-node }	Required Not specified by default.



If b-node is specified for the client, you need to specify no WINS server address.

Configuring the BIMS Server Information for the Client

A DHCP client performs regular software update and backup using configuration files obtained from a branch intelligent management system (BIMS) server. Therefore, the DHCP server needs to offer DHCP clients the BIMS server IP address, port number, shared key from the DHCP address pool.

Follow these steps to configure the BIMS server IP address, port number, and shared key in the DHCP address pool:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter DHCP address pool view	dhcp server ip-pool <i>pool-name</i>	-
Specify the BIMS server IP address, port number, and shared key	bims-server ip <i>ip-address</i> [port <i>port-number</i>] sharekey <i>key</i>	Required Not specified by default.

Configuring Gateways for the Client

DHCP clients that want to access hosts outside the local subnet request gateways to forward data. You can specify gateways in each address pool for clients and the DHCP server will assign gateway addresses while assigning an IP address to the client. Up to eight gateways can be specified in a DHCP address pool.

Follow these steps to configure the gateways in the DHCP address pool:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter DHCP address pool view	dhcp server ip-pool <i>pool-name</i>	-
Specify gateways	gateway-list <i>ip-address&<1-8></i>	Required No gateway is specified by default.

Configuring Option 184 Parameters for the Client with Voice Service

To assign voice calling parameters along with an IP address to DHCP clients with voice service, you need to configure Option 184 on the DHCP server. For information about Option 184, refer to "Option 184" on page 796.

If option 55 in the request from a DHCP client contains option 184, the DHCP server will return parameters specified in option 184 to the client. The client then can initiate a call using parameters in Option 184.

Follow these steps to configure option 184 parameters in the DHCP address pool:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter DHCP address pool view	dhcp server ip-pool <i>pool-name</i>	-
Specify the IP address of the primary network calling processor	voice-config ncp-ip <i>ip-address</i>	Required Not specified by default.
Specify the IP address of the backup network calling processor	voice-config as-ip <i>ip-address</i>	Optional Not specified by default.
Configure the voice VLAN	voice-config voice-vlan <i>vlan-id</i> { disable enable }	Optional Not configured by default.

To do...	Use the command...	Remarks
Specify the failover IP address	voice-config fail-over <i>ip-address dialer-string</i>	Optional No failover IP address is specified by default.



Specify an IP address for the network calling processor before performing other configuration.

Configuring the TFTP Server and Bootfile Name for the Client

This task is to specify the IP address and name of a TFTP server and the bootfile name in the DHCP address pool. The DHCP clients use these parameters to contact the TFTP server, requesting the configuration file used for system initialization, which is called auto-configuration. The request process of the client is described below:

- 1 When a switch starts up without loading any configuration file, the system sets an active interface (such as the VLAN interface of the default VLAN) as the DHCP client to request from the DHCP server parameters such as an IP address and name of a TFTP server, and the bootfile name.
- 2 After getting related parameters, the DHCP client will send a TFTP request to obtain the configuration file from the specified TFTP server for system initialization. If the client cannot get such parameters, it will perform system initialization without loading any configuration file.

To implement auto-configuration, you need to specify the IP address and name of a TFTP server and the bootfile name in the DHCP address pool on the DHCP server, but you do not need to perform any configuration on the DHCP client.

When option 55 in the requesting client message contains parameters of option 66, option 67, or option 150, the DHCP server will return the IP address and name of the specified TFTP server, and bootfile name to the client.

Follow these steps to configure the IP address and name of the TFTP server and the bootfile name in the DHCP address pool:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter DHCP address pool view	dhcp server ip-pool <i>pool-name</i>	-
Specify the TFTP server	tftp-server ip-address <i>ip-address</i>	Optional Not specified by default.
Specify the name of the TFTP server	tftp-server domain-name <i>domain-name</i>	Optional Not specified by default.
Specify the bootfile name	bootfile-name <i>bootfile-name</i>	Optional Not specified by default.

Configuring Self-Defined DHCP Options

By configuring self-defined DHCP options, you can

- Define new DHCP options. New configuration options will come out with DHCP development. To support these new options, you can add them into the attribute list of the DHCP server.

- Define existing DHCP options. Some options have no unified definitions in RFC 2132; however, vendors can define such options as needed. The self-defined DHCP option enables DHCP clients to obtain vendor-specific information.
- Extend existing DHCP options. When the current DHCP options cannot meet the customers' requirements (for example, you cannot use the **dns-list** command to configure more than eight DNS server addresses), you can configure a self defined option for extension.

Follow these steps to configure a self-defined DHCP option in the DHCP address pool:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter DHCP address pool view	dhcp server ip-pool <i>pool-name</i>	-
Configure a self-defined DHCP option	option code { ascii <i>ascii-string</i> hex <i>hex-string</i> &<1-16> ip-address <i>ip-address</i> &<1-8> }	Required No DHCP option is configured by default.

Table 61 Description of common options

Option	Option name	Corresponding command	Command parameter
3	Router Option	gateway-list	ip-address
6	Domain Name Server Option	dns-list	ip-address
15	Domain Name	domain-name	ascii
44	NetBIOS over TCP/IP Name Server Option	nbns-list	ip-address
46	NetBIOS over TCP/IP Node Type Option	netbios-type	hex
51	IP Address Lease Time	expired	hex
58	Renewal (T1) Time Value	expired	hex
59	Rebinding (T2) Time Value	expired	hex
66	TFTP server name	tftp-server	ascii
67	Bootfile name	bootfile-name	ascii
43	Vendor Specific Information	-	hex



CAUTION:

- *Be cautious when configuring self-defined DHCP options because such configuration may affect the DHCP operation process.*
- *When you use self-defined option (Option 51) to configure the IP address lease duration, convert the lease duration into seconds in hexadecimal notation.*

Configuring the DHCP Server Security Functions

This configuration is necessary to secure DHCP services on the DHCP server.

Configuration Prerequisites

Before performing this configuration, complete the following configuration on the DHCP server:

- Enable DHCP
- Configure the DHCP address pool

Enabling Unauthorized DHCP Server Detection

There are unauthorized DHCP servers on networks, which reply DHCP clients with wrong IP addresses.

With this feature enabled, upon receiving a DHCP request, the DHCP server will record the IP address of the DHCP server which assigned an IP address to the DHCP client and the receiving interface. The administrator can use this information to check out any unauthorized DHCP servers.

Follow these steps to enable unauthorized DHCP server detection:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable unauthorized DHCP server detection	dhcp server detect	Required Disabled by default.



With the unauthorized DHCP server detection enabled, the device puts a record once for each DHCP server. The administrator needs to find unauthorized DHCP servers from the log information.

Configuring IP Address Conflict Detection

To avoid IP address conflicts, the DHCP server checks whether the address to be assigned is in use via sending ping packets.

The DHCP server pings the IP address to be assigned using ICMP. If the server gets a response within the specified period, the server will ping another IP address; otherwise, the server will ping the IP addresses once again until the specified number of ping packets are sent. If still no response, the server will assign the IP address to the requesting client (The DHCP client probes the IP address by sending gratuitous ARP packets).

Follow these steps to configure IP address conflict detection:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Specify the number of ping packets	dhcp server ping packets <i>number</i>	Optional One ping packet by default. The value 0 indicates that no ping operation is performed.
Configure a timeout waiting for ping responses	dhcp server ping timeout <i>milliseconds</i>	Optional 500 ms by default. The value 0 indicates that no ping operation is performed.

Configuring the Handling Mode for Option 82

When the DHCP server receives a message with Option 82, if the server is configured to handle Option 82, it will return a response message carrying Option 82 to assign an IP address to the requesting client.

If the server is configured to ignore Option 82, it will assign an IP address to the client without adding Option 82 in the response message.

Configuration prerequisites

Before performing this configuration, complete the following configuration on the DHCP server:

- Enable DHCP
- Configure the DHCP address pool

Configuring the handling mode for Option 82

Follow these steps to enable the DHCP server to handle Option 82:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the server to handle Option 82	dhcp server relay information enable	Optional Enabled by default.



To support Option 82, it is required to perform configuration on both the DHCP server and relay agent (or the device enabled with DHCP Snooping). Refer to “Configuring the DHCP Relay Agent to Support Option 82” on page 818 and “Configuring DHCP Snooping to Support Option 82” on page 828 for related configuration details.

Displaying and Maintaining the DHCP Server

To do...	Use the command...	Remarks
Display information about IP address conflicts	display dhcp server conflict { all ip ip-address }	Available in any view
Display information about lease expiration	display dhcp server expired { all ip ip-address pool [pool-name] }	
Display information about assignable IP addresses	display dhcp server free-ip	
Display IP addresses excluded from dynamic allocation in the DHCP address pool	display dhcp server forbidden-ip	
Display information about bindings	display dhcp server ip-in-use { all ip ip-address pool [pool-name] }	
Display information about DHCP server statistics	display dhcp server statistics	
Display information about the address pool tree organization	display dhcp server tree { all pool [pool-name] }	

To do...	Use the command...	Remarks
Clear information about IP address conflicts	reset dhcp server conflict { all ip <i>ip-address</i> }	Available in user view
Clear information about dynamic bindings	reset dhcp server ip-in-use { all ip <i>ip-address</i> pool [<i>pool-name</i>] }	
Clear information about DHCP server statistics	reset dhcp server statistics	



Using the **save** command does not save DHCP server lease information. Therefore, when the system boots up or the **reset dhcp server ip-in-use** command is executed, no lease information will be available in the configuration file. In this case, the server will deny the request for lease extension from a client and the client needs to request an IP address again.

DHCP Server Configuration Examples

DHCP networking involves two types:

- The DHCP server and client are on the same subnet and exchange messages directly.
- The DHCP server and client are not on the same subnet and they communicate with each other via a DHCP relay agent.

The DHCP server configuration for the two types is the same.

Network requirements

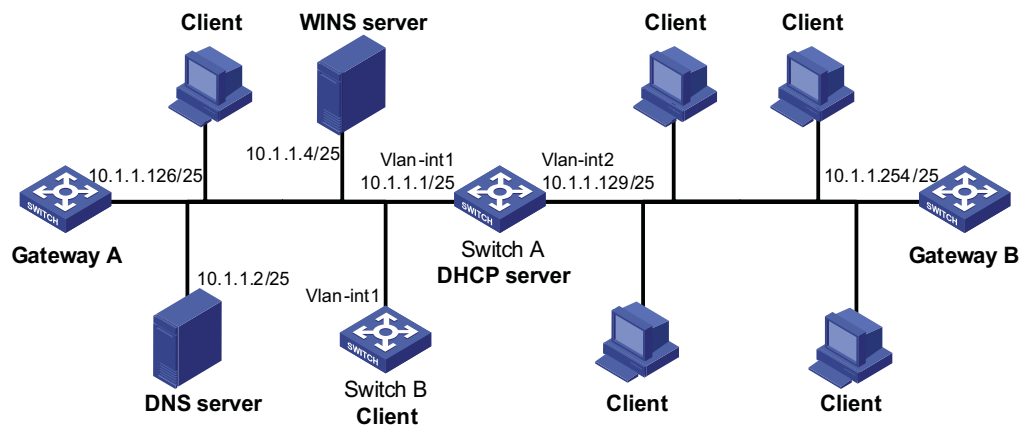
- The DHCP server (Switch A) assigns IP address to clients in subnet 10.1.1.0/24, which is subnetted into 10.1.1.0/25 and 10.1.1.128/25.
- The IP addresses of VLAN-interfaces 1 and 2 on Switch A are 10.1.1.1/25 and 10.1.1.129/25 respectively.
- In the address pool 10.1.1.0/25, the address lease duration is ten days and twelve hours, domain name suffix aabbcc.com, DNS server address 10.1.1.2, gateway 10.1.1.126, and WINS server 10.1.1.4.
- In the address pool 10.1.1.128/25, the address lease duration is five days, domain name suffix aabbcc.com, DNS server address 10.1.1.2, and gateway address 10.1.1.254, and there is no WINS server address.
- The domain name and DNS server address on the subnets 10.1.1.0/25 and 10.1.1.128/25 are the same. Therefore, the domain name suffix and DNS server address can be configured only for the subnet 10.1.1.0/24. The subnet 10.1.1.128/25 can inherit the configuration of the subnet 10.1.1.0/24.



In this example, the number of requesting clients connected to VLAN-interface 1 should be less than 122, and that of clients connected to VLAN-interface 2 less than 124.

Network diagram

Figure 246 DHCP network diagram



Configuration procedure

Specify IP addresses for VLAN interfaces (omitted).

Configure the DHCP server

Enable DHCP.

```
<SwitchA> system-view
[SwitchA] dhcp enable
```

Exclude IP addresses (addresses of the DNS server, WINS server and gateways).

```
[SwitchA] dhcp server forbidden-ip 10.1.1.2
[SwitchA] dhcp server forbidden-ip 10.1.1.4
[SwitchA] dhcp server forbidden-ip 10.1.1.126
[SwitchA] dhcp server forbidden-ip 10.1.1.254
```

Configure DHCP address pool 0 (address range, client domain name suffix, and DNS server address).

```
[SwitchA] dhcp server ip-pool 0
[SwitchA-dhcp-pool-0] network 10.1.1.0 mask 255.255.255.0
[SwitchA-dhcp-pool-0] domain-name aabbcc.com
[SwitchA-dhcp-pool-0] dns-list 10.1.1.2
[SwitchA-dhcp-pool-0] quit
```

Configure DHCP address pool 1 (address range, gateway, lease duration, and WINS server).

```
[SwitchA] dhcp server ip-pool 1
[SwitchA-dhcp-pool-1] network 10.1.1.0 mask 255.255.255.128
[SwitchA-dhcp-pool-1] gateway-list 10.1.1.126
[SwitchA-dhcp-pool-1] expired day 10 hour 12
[SwitchA-dhcp-pool-2] nbns-list 10.1.1.4
[SwitchA-dhcp-pool-1] quit
```

Configure DHCP address pool 2 (address range, gateway, and lease duration).

```
[SwitchA] dhcp server ip-pool 2
[SwitchA-dhcp-pool-2] network 10.1.1.128 mask 255.255.255.128
[SwitchA-dhcp-pool-2] expired day 5
[SwitchA-dhcp-pool-2] gateway-list 10.1.1.254
```

Troubleshooting DHCP Server Configuration

Symptom

A client's IP address obtained from the DHCP server conflicts with another IP address.

Analysis

A host on the subnet may have the same IP address.

Solution

- 1 Disconnect the client's network cable and ping the client's IP address on another host with a long timeout time to check whether there is a host using the same IP address.
- 2 If a ping response is received, the IP address has been manually configured on the host. Execute the **dhcp server forbidden-ip** command on the DHCP server to exclude the IP address from dynamic allocation.
- 3 Connect the client's network cable. Release the IP address and obtain another one on the client. Take WINDOW XP as an example, run **cmd** to enter into DOS window. Type **ipconfig/release** to relinquish the IP address and then **ipconfig/renew** to obtain another IP address.

When configuring the DHCP relay agent, go to these sections for information you are interested in:

- "Introduction to DHCP Relay Agent" on page 813
- "Configuration Task List" on page 815
- "Configuring the DHCP Relay Agent" on page 815
- "Displaying and Maintaining DHCP Relay Agent Configuration" on page 819
- "DHCP Relay Agent Configuration Example" on page 820
- "Troubleshooting DHCP Relay Agent Configuration" on page 821



- *The DHCP relay agent configuration is supported only VLAN interfaces.*
- *DHCP Snooping must be disabled on the DHCP relay agent.*

Introduction to DHCP Relay Agent

Application Environment

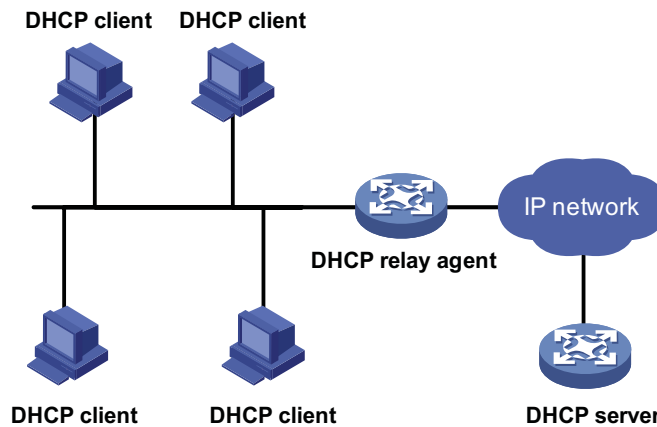
Since DHCP clients request IP addresses via broadcast messages, the DHCP server and clients must be on the same subnet. Therefore, a DHCP server must be available on each subnet. It is not practical.

DHCP relay agent solves the problem. Via a relay agent, DHCP clients communicate with a DHCP server on another subnet to obtain configuration parameters. Thus, DHCP clients on different subnets can contact the same DHCP server for ease of centralized management and cost reduction.

Fundamentals

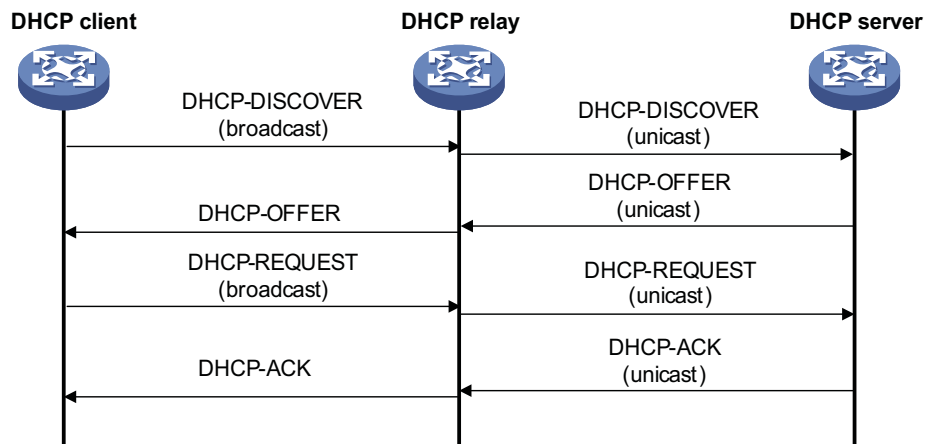
Figure 247 shows a typical application of the DHCP relay agent.

Figure 247 DHCP relay agent application



No matter whether a relay agent exists or not, the DHCP server and client interact with each other in a similar way (see section “Dynamic IP Address Allocation Process” on page 792). The following describes the forwarding process on the DHCP relay agent.

Figure 248 DHCP relay agent work process



As shown in the figure above, the DHCP relay agent works as follows:

- 1 After receiving a DHCP-DISCOVER or DHCP-REQUEST broadcast message from a DHCP client, the DHCP relay agent fills the giaddr field of the message with its IP address and forwards the message to the designated DHCP server in unicast mode.
- 2 Based on the giaddr field, the DHCP server returns an IP address and other configuration parameters to the relay agent, which conveys them to the client.

DHCP Relay Agent Support for Option 82

Option 82 records the location information of the DHCP client. The administrator can locate the DHCP client to further implement security control and accounting. For more information, refer to “Relay agent option (Option 82)” on page 795.

If the DHCP relay agent supports Option 82, it will handle a client’s request according to the contents defined in Option 82, if any. The handling strategies are described in the table below.

If a reply returned by the DHCP server contains Option 82, the DHCP relay agent will remove the Option 82 before forwarding the reply to the client.

If a client's requesting message has...	Handling strategy	Padding format	The DHCP relay agent will...
Option 82	Drop	Random	Drop the message.
	Keep	Random	Forward the message without changing Option 82.
	Replace	normal	Forward the message after replacing the original Option 82 with the Option 82 padded in normal format.
verbose		Forward the message after replacing the original Option 82 with the Option 82 padded in verbose format.	
no Option 82	-	normal	Forward the message after adding the Option 82 padded in normal format.
	-	verbose	Forward the message after adding the Option 82 padded in verbose format.

Configuration Task List

Complete the following tasks to configure the DHCP relay agent:

Task	Remarks
"Enabling DHCP" on page 815	Required
"Enabling the DHCP Relay Agent on an Interface" on page 815	Required
"Correlating a DHCP Server Group with a Relay Agent Interface" on page 816	Required
"Configuring the DHCP Relay Agent to Send a DHCP-Release Request" on page 816	Optional
"Configuring the DHCP Relay Agent Security Functions" on page 817	Optional
"Configuring the DHCP Relay Agent to Support Option 82" on page 818	Optional

Configuring the DHCP Relay Agent

Enabling DHCP Enable DHCP before performing other DHCP-related configurations.

Follow these steps to enable DHCP:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable DHCP	dhcp enable	Required Disabled by default.

Enabling the DHCP Relay Agent on an Interface

With this task completed, upon receiving a DHCP request from the enabled interface, the relay agent will forward the request to a DHCP server for address allocation.

Follow these steps to enable the DHCP relay agent on an interface:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable the DHCP relay agent on the current interface	dhcp select relay	Required With DHCP enabled, interfaces work in the DHCP server mode.



If the DHCP client obtains an IP address via the DHCP relay agent, the address pool of the subnet which the IP address of the DHCP relay agent belongs to must be configured on the DHCP server. Otherwise, the DHCP client cannot obtain a correct IP address.

Correlating a DHCP Server Group with a Relay Agent Interface

To improve reliability, you can specify several DHCP servers as a group on the DHCP relay agent and correlate a relay agent interface with the server group. When the interface receives requesting messages from clients, the relay agent will forward them to all the DHCP servers of the group.

Follow these steps to correlate a DHCP server group with a relay agent interface:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a DHCP server group and add a server into the group	dhcp relay server-group <i>group-id</i> ip <i>ip-address</i>	Required Not created by default.
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Correlate the DHCP server group with the current interface	dhcp relay server-select <i>group-id</i>	Required By default, no interface is correlated with any DHCP server group.



- You can specify at most twenty DHCP server groups on the relay agent and at most eight DHCP server addresses for each DHCP server group.
- The IP addresses of DHCP servers and those of relay agent's interfaces cannot be on the same subnet. Otherwise, the client cannot obtain an IP address.
- A DHCP server group can correlate with one or multiple DHCP relay agent interfaces, while a relay agent interface can only correlate with one DHCP server group. Using the **dhcp relay server-select** command repeatedly overwrites the previous configuration. However, if the specified DHCP server group does not exist, the interface still uses the previous correlation.
- The *group-id* in the **dhcp relay server-select** command was specified by the **dhcp relay server-group** command.

Configuring the DHCP Relay Agent to Send a DHCP-Release Request

Sometimes, you need to release a client's IP address manually on the DHCP relay agent. With this task completed, the DHCP relay agent can actively send a DHCP-RELEASE request that contains the client's IP address to be released. Upon

receiving the DHCP-RELEASE request, the DHCP server then releases the IP address for the client.

Follow these steps to configure the DHCP relay agent in system view to send a DHCP-RELEASE request:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the DHCP relay agent to send a DHCP-RELEASE request	dhcp relay release ip <i>client-ip</i>	Required

Configuring the DHCP Relay Agent Security Functions

Creating static bindings and enabling IP address check

The DHCP relay agent can dynamically record clients' IP-to-MAC bindings after clients get IP addresses. It also supports static bindings, which means you can manually configure IP-to-MAC bindings on the DHCP relay agent, so that users can access external network using fixed IP addresses.

For avoidance of invalid IP address configuration, you can configure the DHCP relay agent to check whether a requesting client's IP and MAC addresses match a binding on it (both dynamic and static bindings). If not, the client cannot access outside networks via the DHCP relay agent.

Follow these steps to create a static binding and enable IP address check:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a static binding	dhcp relay security static <i>ip-address</i> <i>mac-address</i> [interface <i>interface-type</i> <i>interface-number</i>]	Optional No static binding is created by default.
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable invalid IP address check	dhcp relay address-check { disable enable }	Required Disabled by default.



- The **dhcp relay address-check enable** command is independent of other commands of the DHCP relay agent. That is, the invalid address check takes effect when this command is executed, regardless of whether other commands are used.
- You are recommended to configure IP address check on the interface enabled with the DHCP relay agent; otherwise, the valid DHCP clients may not be capable of accessing networks.
- When using the **dhcp relay security static** command to bind a VLAN interface to a static binding entry, make sure that the VLAN interface is configured as a DHCP relay agent; otherwise, address entry conflicts may occur.

Configuring dynamic binding update interval

Via the DHCP relay agent, a DHCP client sends a DHCP-RELEASE unicast message to the DHCP server to relinquish its IP address. In this case the DHCP relay agent

simply conveys the message to the DHCP server, thus it does not remove the IP address from its bindings. To solve this, the DHCP relay agent can update dynamic bindings at a specified interval.

The DHCP relay agent uses the IP address of a client and the MAC address of the DHCP relay interface to regularly send a DHCP-REQUEST message to the DHCP server.

- If the server returns a DHCP-ACK message or does not return any message within a specified interval, which means the IP address is assignable now, the DHCP relay agent will update its bindings by aging out the binding entry of the IP address.
- If the server returns a DHCP-NAK message, which means the IP address is still in use, the relay agent will not age it out.

Follow these steps to configure dynamic binding update interval:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure binding update interval	dhcp relay security tracker { <i>interval</i> auto }	Optional auto by default. (auto interval is calculated by the relay agent according to the number of bindings.)

Enabling unauthorized DHCP servers detection

There are unauthorized DHCP servers on networks, which reply DHCP clients with wrong IP addresses.

With this feature enabled, upon receiving a DHCP request, the DHCP relay agent will record the IP address of the DHCP server which assigned an IP address to the DHCP client and the receiving interface. The administrator can use this information to check out any DHCP unauthorized servers.

Follow these steps to enable unauthorized DHCP server detection:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable unauthorized DHCP server detection	dhcp relay server-detect	Required Disabled by default.



With the unauthorized DHCP server detection enabled, the device puts a record once for each DHCP server. The administrator needs to find unauthorized DHCP servers from the log information. After the recorded information of a DHCP server is cleared, a new record will be put for the DHCP server.

Configuring the DHCP Relay Agent to Support Option 82

Prerequisites

You need to complete the following tasks before configuring the DHCP relay agent to support Option 82.

- Enabling DHCP

- Enabling the DHCP relay agent on the specified interface
- Correlating a DHCP server group with relay agent interfaces

Configuring the DHCP relay agent to support Option 82

Follow these steps to configure the DHCP relay agent to support Option 82:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable the relay agent to support Option 82	dhcp relay information enable	Required Disabled by default.
Configure the handling strategy for requesting messages containing Option 82	dhcp relay information strategy { drop keep replace }	Optional replace by default.
Configure the padding format for Option 82	dhcp relay information format { normal verbose [node-identifier { mac sysname user-defined <i>node-identifier</i>] } }	Optional normal by default.



- *To support Option 82, it is required to perform related configuration on both the DHCP server and relay agent. Refer to “Configuring the Handling Mode for Option 82” on page 808 for DHCP server configuration of this kind.*
- *If the handling strategy of the DHCP relay agent is configured as **replace**, you need to configure a padding format for Option 82. If the handling strategy is **keep** or **drop**, you need not configure any padding format.*
- *If sub-option 1 (node identifier) of Option 82 is padded with the device name (sysname) of a node, the device name must contain no spaces. Otherwise, the DHCP relay agent will drop the message.*

Displaying and Maintaining DHCP Relay Agent Configuration

To do...	Use the command...	Remarks
Display information about DHCP server groups correlated to a specified or all interfaces	display dhcp relay { all interface <i>interface-type interface-number</i> }	Available in any view
Display information about bindings of DHCP relay agents	display dhcp relay security [<i>ip-address</i> dynamic static]	Available in any view
Display statistics information about bindings of DHCP relay agents	display dhcp relay security statistics	Available in any view
Display information about the refreshing interval for entries of dynamic IP-to-MAC bindings	display dhcp relay security tracker	Available in any view
Display information about the configuration of a specified or all DHCP server groups	display dhcp relay server-group { <i>group-id</i> all }	Available in any view
Display packet statistics on relay agent	display dhcp relay statistics [server-group { <i>group-id</i> all }]	Available in user view
Clear packet statistics from relay agent	reset dhcp relay statistics [server-group <i>group-id</i>]	Available in user view

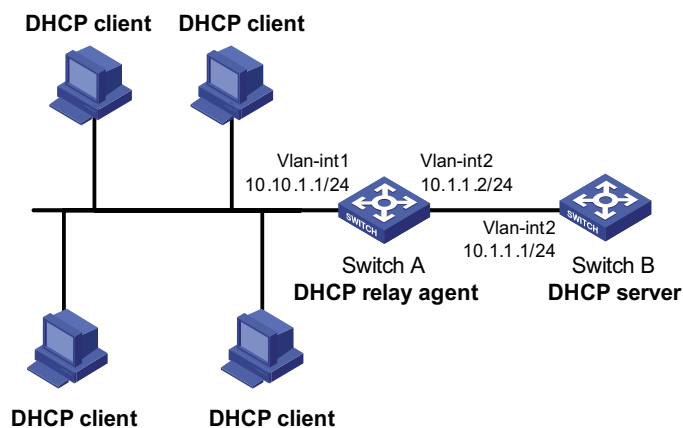
DHCP Relay Agent Configuration Example

Network requirements

VLAN-interface 1 on the DHCP relay agent (Switch A) connects to the network where DHCP clients reside. The IP address of VLAN-interface 1 is 10.10.1.1/24 and IP address of VLAN-interface 2 is 10.1.1.2/24 that communicates with the DHCP server 10.1.1.1/24. As shown in the figure below, Switch A forwards messages between DHCP clients and the DHCP server.

Network diagram

Figure 249 Network diagram for DHCP relay agent



Configuration procedure

Enable DHCP.

```
<SwitchA> system-view
[SwitchA] dhcp enable
```

Enable the DHCP relay agent on VLAN-interface 1.

```
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] dhcp select relay
[SwitchA-Vlan-interface1] quit
```

Configure DHCP server group 1 with the DHCP server 10.1.1.1, and correlate the DHCP server group 1 with VLAN-interface 1.

```
[SwitchA] dhcp relay server-group 1 ip 10.1.1.1
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] dhcp relay server-select 1
```



- *Performing the configuration on the DHCP server is also required to guarantee the client-server communication via the relay agent. Refer to “DHCP Server Configuration Examples” on page 809 for DHCP server configuration information.*
- *If the DHCP relay agent and server are on different subnets, routes in between must be reachable.*

**Troubleshooting DHCP
Relay Agent
Configuration****Symptom**

DHCP clients cannot obtain any configuration parameters via the DHCP relay agent.

Analysis

Some problems may occur with the DHCP relay agent or server configuration. Enable debugging and execute the **display** command on the DHCP relay agent to view the debugging information and interface state information for locating the problem.


Solution

Check that:

- The DHCP is enabled on the DHCP server and relay agent.
- The address pool on the same subnet where DHCP clients reside is available on the DHCP server.
- The routes between the DHCP server and DHCP relay agent are reachable.
- The relay agent interface connected to DHCP clients is correlated with correct DHCP server group and IP addresses for the group members are correct.

DHCP CLIENT CONFIGURATION

When configuring the DHCP client, go to these sections for information you are interested in:

- “Introduction to DHCP Client” on page 823
 - “Enabling the DHCP Client on an Interface” on page 823
 - “Displaying and Maintaining the DHCP Client” on page 824
 - “DHCP Client Configuration Example” on page 824
-  ■ *The DHCP client configuration is supported only on VLAN interfaces.*
- *When multiple VLAN interfaces with the same MAC address use DHCP for IP address acquisition via a relay agent, the DHCP server cannot be a Windows 2000 Server or Windows 2003 Server.*
- *You are not recommended to enable both the DHCP client and the DHCP Snooping on the same device. Otherwise, DHCP Snooping entries may fail to be generated, or the DHCP client may fail to obtain an IP address.*

Introduction to DHCP Client

With the DHCP client enabled on an interface, the interface will use DHCP to obtain configuration parameters such as an IP address from the DHCP server.

For the Switch 4800G (operating as DHCP clients), the vendor and device information contained in Option 60 of DHCP requests is not configurable; instead, it is determined by the application program of the switches. Refer to Table 62 for different information added in Option 60 based on device models.

Table 62 Description on the vendor and device information in Option 60

Device Model	Vendor and device information
3CRS48G-24-91	3Com Switch 4800G 24-Port
3CRS48G-24P-91	3Com Switch 4800G PWR 24-Port
3CRS48G-48-91	3Com Switch 4800G 48-Port
3CRS48G-48P-91	3Com Switch 4800G PWR 48-Port
3CRS48G-24S-91	3Com Switch 4800G 24-Port SFP

Enabling the DHCP Client on an Interface

Follow these steps to enable the DHCP client on an interface:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-

To do...	Use the command...	Remarks
Enable the DHCP client on the interface	ip address dhcp-alloc [client-identifier mac <i>interface-type</i> <i>interface-number</i>]	Required Disabled by default.



- *An interface can be configured to acquire an IP address in multiple ways, but these ways are exclusive. The latest configuration will overwrite the previous configuration.*
- *After the DHCP client is enabled on an interface, no secondary IP address is configurable for the interface.*
- *If the IP address assigned by the DHCP server shares a network segment with the IP addresses of other interfaces on the device, the DHCP client enabled interface will not request any IP address of the DHCP server unless the conflicted IP address is manually deleted and the interface is made UP again by first executing the **shutdown** command and then the **undo shutdown** command or the DHCP client is enabled on the interface by executing the **undo ip address dhcp-alloc** and **ip address dhcp-alloc** commands in sequence.*

Displaying and Maintaining the DHCP Client

To do...	Use the command...	Remarks
Display specified configuration information	display dhcp client [verbose] [interface <i>interface-type</i> <i>interface-number</i>]	Available in any view

DHCP Client Configuration Example

Network requirements

On a LAN, Switch B contacts the DHCP server via VLAN-interface 1 to obtain an IP address.

Network diagram

See Figure 246.

Configuration procedure

The following is the configuration on Switch B shown in Figure 246.

Enable the DHCP client on VLAN-interface 1.


```
<SwitchB> system-view
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ip address dhcp-alloc
```



To implement the DHCP client-server model, you need to perform related configuration on the DHCP server. For details, refer to “DHCP Server Configuration Examples” on page 809.

DHCP SNOOPING CONFIGURATION

When configuring DHCP snooping, go to these sections for information you are interested in:

- “DHCP Snooping Overview” on page 825
 - “Configuring DHCP Snooping Basic Functions” on page 828
 - “Configuring DHCP Snooping to Support Option 82” on page 828
 - “Displaying and Maintaining DHCP Snooping” on page 829
 - “DHCP Snooping Configuration Example” on page 829
-  *DHCP Snooping supports no link aggregation. If an Ethernet port is added into an aggregation group, DHCP Snooping configuration on it will not take effect. When the port is removed from the group, DHCP Snooping can take effect.*
- *The DHCP snooping enabled device does not work if it is between the DHCP relay agent and DHCP server, and it can work when it is between the DHCP client and relay agent or between the DHCP client and server.*
 - *The DHCP Snooping enabled device cannot be a DHCP server or DHCP relay agent.*
 - *You are not recommended to enable the DHCP client, BOOTP client, and DHCP Snooping on the same device. Otherwise, DHCP Snooping entries may fail to be generated, or the BOOTP client/DHCP client may fail to obtain an IP address.*

DHCP Snooping Overview

Function of DHCP Snooping

As a DHCP security feature, DHCP snooping can implement the following:

Recording IP-to-MAC mappings of DHCP clients

For security sake, a network administrator needs to record the mapping between a client's IP address obtained from the DHCP server and the client's MAC address. DHCP snooping can meet the need.

DHCP snooping records clients' MAC and IP addresses by reading their DHCP-REQUEST and DHCP-ACK messages from trusted ports. The network administrator can check out which IP addresses are assigned to the DHCP clients with the **display dhcp-snooping** command.

Ensuring DHCP clients to obtain IP addresses from valid DHCP servers

If there is an unauthorized DHCP server on a network, the DHCP clients may obtain invalid IP addresses. With DHCP snooping, the ports of a device can be

configured as trusted or untrusted, ensuring the clients to obtain IP addresses from authorized DHCP servers.

- Trusted: A trusted port forwards DHCP messages, ensuring that DHCP clients can obtain valid IP addresses.
- Untrusted: The DHCP-ACK or DHCP-OFFER packets received from an untrusted port are discarded, preventing DHCP clients from receiving invalid IP addresses.

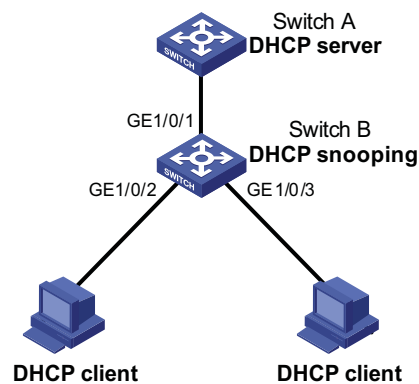
Application Environment of Trusted Ports

Configuring a trusted port connected with a DHCP server

A port that is connected with a DHCP server directly or indirectly should be configured as a trusted port, so that the DHCP snooping device can forward reply messages from the DHCP server, ensuring the DHCP clients to obtain IP addresses from the authorized DHCP server.

As shown in Figure 250, GE1/0/1 on Switch B is connected with Switch A (a DHCP server). GE1/0/1 should be configured as a trusted port, so that it can forward replies from Switch A.

Figure 250 Configure a trusted port connected with the DHCP sever

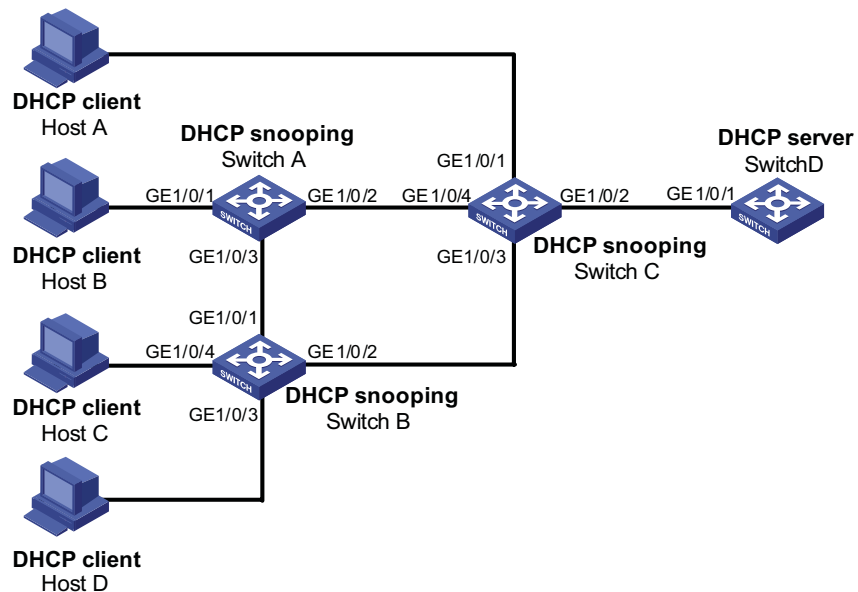


Configuring trusted ports in a cascaded network

In a cascaded network involving multiple DHCP snooping devices, the ports connected to other DHCP snooping devices should be configured as trusted ports.

To save system resources, you can disable the trusted ports, which are indirectly connected with DHCP clients, from recording clients' IP-to-MAC bindings.

As shown in Figure 251, Switch A, Switch B, and Switch C are DHCP snooping devices. GE1/0/2 and GE1/0/3 on Switch A, GE1/0/1 and GE1/0/2 on Switch B, and GE1/0/2, GE1/0/3, and GE1/0/4 on Switch C are configured as trusted ports. Disable the trusted ports, GE1/0/3 on Switch A, GE1/0/1 on Switch B, GE1/0/3 and GE1/0/4 on Switch C, which are not directly connected to DHCP clients, from recording client's IP-to-MAC bindings.

Figure 251 Configure trusted ports in a cascaded network

DHCP Snooping Support for Option 82

Option 82 records the location information of the DHCP client. The administrator can locate the DHCP client to further implement security control and accounting. For more information, refer to “Relay agent option (Option 82)” on page 795.

If DHCP snooping supports Option 82, it will handle a client’s request according to the contents defined in Option 82, if any. The handling strategies are described in the table below.

If a reply returned by the DHCP server contains Option 82, the DHCP snooping device will remove the Option 82 before forwarding the reply to the client. If the reply contains no Option 82, it forwards it directly.

If a client’s requesting message has...	Handling strategy	Padding format	The DHCP snooping device will...
Option 82	Drop	Random	Drop the message.
	Keep	Random	Forward the message without changing Option 82.
	Replace	normal	Forward the message after replacing the original Option 82 with the Option 82 padded in normal format.
verbose		Forward the message after replacing the original Option 82 with the Option 82 padded in verbose format.	
no Option 82	-	normal	Forward the message after adding the Option 82 padded in normal format.
	-	verbose	Forward the message after adding the Option 82 padded in verbose format.



The handling strategy and padding format for Option 82 on the DHCP-Snooping device are the same as those on the relay agent.

Configuring DHCP Snooping Basic Functions

Follow these steps to configure DHCP snooping basic functions:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable DHCP snooping	dhcp-snooping	Required Disabled by default.
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	-
Specify the port as trusted	dhcp-snooping trust [no-user-binding]	Required Untrusted by default.



- You need to specify the ports connected to the valid DHCP servers as trusted to ensure that DHCP clients can obtain valid IP addresses. The trusted port and the port connected to the DHCP client must be in the same VLAN.
- You are not recommended to configure both the DHCP snooping and selective Q-in-Q function on the switch, which may result in the DHCP snooping to function abnormally.

Configuring DHCP Snooping to Support Option 82

Prerequisites

You need to enable the DHCP Snooping function before configuring DHCP Snooping to support Option 82.

Configuring DHCP Snooping to Support Option 82

Follow these steps to configure DHCP snooping to support Option 82:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable DHCP Snooping to support Option 82	dhcp-snooping information enable	Required Disabled by default.
Configure the handling strategy for requesting messages containing Option 82	dhcp-snooping information strategy { drop keep replace }	Optional replace by default.
Configure the padding format for Option 82	dhcp-snooping information format { normal verbose [node-identifier { mac sysname user-defined <i>node-identifier</i> }] }	Optional normal by default.



- To support Option 82, it is required to perform related configuration on both the DHCP server and the device enabled with DHCP Snooping. Refer to "Configuring the Handling Mode for Option 82" on page 808 for DHCP server configuration of this kind.

- If the handling strategy of the DHCP-Snooping-enabled device is configured as **replace**, you need to configure a padding format for Option 82. If the handling strategy is **keep** or **drop**, you need not configure any padding format.
- If the Option 82 is padded with the device name (sysname) of a node, the device name must contain no spaces. Otherwise, the DHCP-Snooping-enabled device will drop the message.

Displaying and Maintaining DHCP Snooping

To do...	Use the command...	Remarks
Display DHCP snooping address binding information	display dhcp-snooping	Available in any view
Display information about trusted ports	display dhcp-snooping trust	
Clear DHCP snooping address binding information	reset dhcp-snooping { all ip ip-address }	Available in user view

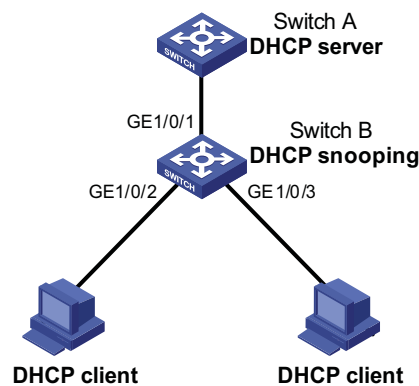
DHCP Snooping Configuration Example

Network requirements

- Switch B is connected to a DHCP server through GigabitEthernet 1/0/1, and to two DHCP clients through GigabitEthernet 1/0/2 and GigabitEthernet 1/0/3.
- GigabitEthernet 1/0/1 forwards DHCP server responses while the other two do not.
- Switch B records clients' IP-to-MAC address bindings in DHCP-REQUEST messages and DHCP-ACK messages received from trusted ports.
- Switch B supports Option 82. After receiving a DHCP request from the client, Switch B adds Option 82 padded in verbose format to the request message and forwards the message to the DHCP server.

Network diagram

Figure 252 Network diagram for DHCP snooping configuration



Configuration procedure

Enable DHCP snooping.

```

<SwitchB> system-view
[SwitchB] dhcp-snooping
  
```

Specify GigabitEthernet 1/0/1 as trusted port.

```
[SwitchB] interface gigabitethernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] dhcp-snooping trust
[SwitchB-GigabitEthernet1/0/1] quit
```

Configure DHCP Snooping to support Option 82 on GigabitEthernet 1/0/2.

```
[SwitchB] interface gigabitethernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] dhcp-snooping information enable
```

Configure the padding format to verbose for Option 82 on GigabitEthernet 1/0/2.

```
[SwitchB-GigabitEthernet1/0/2] dhcp-snooping information format verbose node-identifier sysname
[SwitchB-GigabitEthernet1/0/2] quit
```

Configure DHCP Snooping to support Option 82 on GigabitEthernet 1/0/3.

```
[SwitchB] interface gigabitethernet 1/0/3
[SwitchB-GigabitEthernet1/0/3] dhcp-snooping information enable
```

Configure the padding format to verbose for Option 82 on GigabitEthernet 1/0/3.

```
[SwitchB-GigabitEthernet1/0/3] dhcp-snooping information format verbose node-identifier sysname
```

61

BOOTP CLIENT CONFIGURATION

While configuring a BOOTP client, go to these sections for information you are interested in:

- "Introduction to BOOTP Client" on page 831
- "Configuring an Interface to Dynamically Obtain an IP Address Through BOOTP" on page 832
- "Displaying and Maintaining BOOTP Client Configuration" on page 832



- *BOOTP client configuration only applies to VLAN interfaces.*
- *If several VLAN interfaces sharing the same MAC address obtain IP addresses through a BOOTP relay agent, the BOOTP server cannot be a Windows 2000 Server or Windows 2003 Server.*
- *You are not recommended to enable both the DHCP client and the DHCP Snooping on the same device. Otherwise, DHCP Snooping entries may fail to be generated, or the BOOTP client may fail to obtain an IP address.*

Introduction to BOOTP Client

This section covers these topics:

- "BOOTP Application" on page 831
- "Obtaining an IP Address Dynamically" on page 832
- "Protocols and Standards" on page 832

BOOTP Application

After you specify an interface of a device as a BOOTP client, the interface can use BOOTP to get information (such as IP address) from the BOOTP server, which simplifies your configuration.

Before using BOOTP, an administrator needs to configure a BOOTP parameter file for each BOOTP client on the BOOTP server. The parameter file contains information such as MAC address and IP address of a BOOTP client. When a BOOTP client originates a request to the BOOTP server, the BOOTP server will search for the BOOTP parameter file and return the corresponding configuration information.

Because you need to configure a parameter file for each client on the BOOTP server, BOOTP usually runs under a relatively stable environment. If the network changes frequently, DHCP is applicable.



Because a DHCP server can interact with a BOOTP client, you can use the DHCP server to configure an IP address for the BOOTP client, without any BOOTP server.

Obtaining an IP Address Dynamically



A DHCP server can take the place of the BOOTP server in the following dynamic IP address acquisition.

A BOOTP client dynamically obtains an IP address from a BOOTP server in the following way:

- 1 The BOOTP client broadcasts a BOOTP request, which contains its own MAC address.
- 2 The BOOTP server receives the request and searches the configuration file for the corresponding IP address according to the MAC address of the BOOTP client. The BOOTP server then returns a BOOTP response to the BOOTP client.
- 3 The BOOTP client obtains the IP address from the received the response.

Protocols and Standards

Some protocols and standards related to BOOTP include:

- RFC 951: Bootstrap Protocol (BOOTP)
- RFC 2132: DHCP Options and BOOTP Vendor Extensions
- RFC 1542: Clarifications and Extensions for the Bootstrap Protocol

Configuring an Interface to Dynamically Obtain an IP Address Through BOOTP

Follow these steps to configure an interface to dynamically obtain an IP address:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure an interface to dynamically obtain IP address through BOOTP	ip address bootp-alloc	Required By default, an interface does not use BOOTP to obtain an IP address.

Displaying and Maintaining BOOTP Client Configuration

To do...	Use the command...	Remarks
Display related information on a BOOTP client	display bootp client [interface <i>interface-type</i> <i>interface-number</i>]	Available in any view

BOOTP Client Configuration Example

Network requirement

Switch B's port belonging to VLAN 1 is connected to the LAN. VLAN-interface 1 obtains an IP address from the DHCP server by using BOOTP.

Network diagram

See Figure 246.

Configuration procedure

The following describes only the configuration on Switch B serving as a client.

Configure VLAN-interface 1 to dynamically obtain an IP address from the DHCP server.

```
<SwitchB> system-view
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ip address bootp-alloc
```



To make the BOOTP client to obtain an IP address from the DHCP server, you need to perform additional configurations on the DHCP server. For details, refer to “DHCP Server Configuration Examples” on page 809.

In order to filter traffic, network devices use sets of rules, called access control lists (ACLs), to identify and handle packets.

When configuring ACLs, go to these chapters for information you are interested in:

- "ACL Overview" on page 835
- "IPv4 ACL Configuration" on page 841
- "IPv6 ACL Configuration" on page 851



Unless otherwise stated, ACLs refer to both IPv4 ACLs and IPv6 ACLs throughout this document.

Introduction to ACL

Introduction As network scale and network traffic are increasingly growing, network security and bandwidth allocation become more and more critical to network management. Packet filtering can be used to efficiently prevent illegal users from accessing networks and to control network traffic and save network resources. Access control lists (ACL) are often used to filter packets with configured matching rules.

ACLs are sets of rules (or sets of permit or deny statements) that decide what packets can pass and what should be rejected based on matching criteria such as source MAC address, destination MAC address, source IP address, destination IP address, and port number.

Application of ACLs on the Switch

The switch supports two ACL application modes:

- Hardware-based application: An ACL is assigned to a piece of hardware. For example, an ACL can be referenced by QoS for traffic classification. Note that when an ACL is referenced to implement QoS, the actions defined in the ACL rules, deny or permit, do not take effect; actions to be taken on packets matching the ACL depend on the traffic behavior definition in QoS. For details about traffic behavior, refer to "Traffic Classification, TP, and LR Configuration" on page 861.
- Software-based application: An ACL is referenced by a piece of upper layer software. For example, an ACL can be referenced to configure login user control behavior, thus controlling Telnet, SNMP and Web users. Note that when an ACL is reference by the upper layer software, actions to be taken on packets matching the ACL depend on those defined by the ACL rules. For details about login user control, refer to "Controlling Login Users" on page 75.



- When an ACL is assigned to a piece of hardware and referenced by a QoS policy for traffic classification, the switch does not take action according to the traffic behavior definition on a packet that does not match the ACL.
- When an ACL is referenced by a piece of software to control Telnet, SNMP, and Web login users, the switch denies all packets that do not match the ACL.

Introduction to IPv4 ACL

This section covers these topics:

- “IPv4 ACL Classification” on page 836
- “IPv4 ACL Naming” on page 836
- “IPv4 ACL Match Order” on page 836
- “IPv4 ACL Step” on page 837
- “Effective Period of an IPv4 ACL” on page 838
- “IP Fragments Filtering with IPv4 ACL” on page 838

IPv4 ACL Classification

IPv4 ACLs, identified by ACL numbers, fall into four categories, as shown in Table 63.

Table 63 IPv4 ACL categories

Category	ACL number	Matching criteria
Basic IPv4 ACL	2000 to 2999	Source IP address
Advanced IPv4 ACL	3000 to 3999	Source IP address, destination IP address, protocol carried on IP, and other Layer 3 or Layer 4 protocol header information
Ethernet frame header ACL	4000 to 4999	Layer 2 protocol header fields such as source MAC address, destination MAC address, 802.1p priority, and link layer protocol type

IPv4 ACL Naming

When creating an IPv4 ACL, you can specify a unique name for it. Afterwards, you can identify the ACL by its name.

An IPv4 ACL can have only one name. Whether to specify a name for an ACL is up to you. After creating an ACL, you cannot specify a name for it, nor can you change or remove the name of the ACL.



The name of an IPv4 ACL must be unique among IPv4 ACLs. However, an IPv4 ACL and an IPv6 ACL can share the same name.

IPv4 ACL Match Order

An ACL consists of multiple rules, each of which specifies different matching criteria. These criteria may have overlapping or conflicting parts. This is where the order in which a packet is matched against the rules comes to rescue.

Two match orders are available for IPv4 ACLs:

- **config**: where packets are compared against ACL rules in the order in which they are configured.
- **auto**: where depth-first match is performed. The term depth-first match has different meanings for different types of ACLs.

Depth-first match for a basic IPv4 ACL

The following shows how your switch performs depth-first match in a basic IPv4 ACL:

- 1 Sort rules by source IP address wildcard first and compare packets against the rule configured with more zeros in the source IP address wildcard prior to other rules.
- 2 If two rules are present with the same number of zeros in their source IP address wildcards, compare packets against the rule configured first prior to the other.

Depth-first match for an advanced IPv4 ACL

The following shows how your switch performs depth-first match in an advanced IPv4 ACL:

- 1 Sort rules by protocol range and compare packets against the rule with the protocol carried on IP specified prior to the other.
- 2 If the protocol ranges are the same, look at source IP address wildcard. Then, compare packets against the rule configured with more zeros in the source IP address wildcard prior to the other.
- 3 If the numbers of zeros in the source IP address wildcards are the same, look at the destination IP address wildcards. Then, compare packets against the rule configured with more zeros in the destination IP address wildcard prior to the other.
- 4 If the numbers of zeros in the destination IP address wildcards are the same, look at the Layer 4 port number (TCP/UDP port number). Then compare packets against the rule configured with the lower port number prior to the other.
- 5 If the port numbers are the same, compare packets against the rule configured first prior to the other.

Depth-first match for an Ethernet frame header ACL

The following shows how your switch performs depth-first match in an Ethernet frame header ACL:

- 1 Sort rules by source MAC address mask first and compare packets against the rule configured with more ones in the source MAC address mask prior to other rules.
- 2 If two rules are present with the same number of ones in their source MAC address masks, look at the destination MAC address masks. Then, compare packets against the rule configured with more ones in the destination MAC address mask prior to the other.
- 3 If the numbers of ones in the destination MAC address masks are the same, the one configured first is compared prior to the other.

The comparison of a packet against an ACL stops once a match is found. The packet is then processed as per the rule.

IPv4 ACL Step Meaning of the step

When defining rules in an IPv4 ACL, you do not necessarily assign them numbers; the system can do this automatically, and the step defines the increment between two neighboring numbers. For example, with a step of 5, rules are automatically numbered 0, 5, 10, 15, and so on. By default, the step is 5.

Whenever the step changes, the rules are renumbered. Continuing with the above example, if you change the step from 5 to 2, the rules are renumbered 0, 2, 4, 6, and so on.

Benefits of using the step

With the step and rule numbering/renumbering mechanism, you do not need to assign rules numbers when defining them. The system will assign a newly defined rule a number that is the smallest multiple of the step bigger than the currently biggest number. For example, with a step of five, if the biggest number is currently 28, the newly defined rule will get a number of 30. If the ACL has no rule defined already, the first defined rule will get a number of 0.

Another benefit of using the step is that it allows you to insert new rules between existing ones as needed. For example, after creating four rules numbered 0, 5, 10, and 15 in an ACL with a step of five, you can insert a rule numbered 1.

Effective Period of an IPv4 ACL

You can control when a rule can take effect by referencing a time range in the rule.

A referenced time range can be one that has not been created yet. The rule, however, can take effect only after the time range is defined and comes active.

IP Fragments Filtering with IPv4 ACL

Traditional packet filtering performs match operation on, rather than all IP fragments, the first ones only. All subsequent non-first fragments are handled in the way the first fragments are handled. This causes security risk as attackers may fabricate non-first fragments to attack your network.

As for the configuration of a rule of an IPv4 ACL, the **fragment** keyword specifies that the rule applies to non-first fragment packets only, and does not apply to non-fragment packets or the first fragment packets. ACL rules that do not contain this keyword is applicable to both non-fragment packets and fragment packets.

Introduction to IPv6 ACL

This section covers these topics:

- "IPv6 ACL Classification" on page 838
- "IPv6 ACL Naming" on page 839
- "IPv6 ACL Match Order" on page 839
- "IPv6 ACL Step" on page 840
- "Effective Period of an IPv6 ACL" on page 840

IPv6 ACL Classification

IPv6 ACLs, identified by ACL numbers, fall into three categories, as show in Table 64.

Table 64 IPv6 ACL categories

Category	ACL number	Matching criteria
Basic IPv6 ACL	2000 to 2999	Source IPv6 address
Advanced IPv6 ACL	3000 to 3999	Source IPv6 address, destination IPv6 address, protocol carried on IPv6, and other Layer 3 or Layer 4 protocol header fields

IPv6 ACL Naming When creating an IPv6 ACL, you can specify a unique name for it. Afterwards, you can identify the IPv6 ACL by its name.

An IPv6 ACL can have only one name. Whether to specify a name for an ACL is up to you. After creating an ACL, you cannot specify a name for it, nor can you change or remove the name of the ACL.



The name of an IPv6 ACL must be unique among IPv6 ACLs. However, an IPv6 ACL and an IPv4 ACL can share the same name.

IPv6 ACL Match Order Similar to IPv4 ACLs, IPv6 ACLs are sequential collections of rules defined with different matching parameters. The order in which a packet is matched against the rules in an IPv6 ACL may affect how the packet is handled.

Like in IPv4 ACLs, the following two match orders are available in IPv6 ACLs:

- **config**: where rules are compared against in the order in which they are configured.
- **auto**: where depth-first match is performed.

Depth-first match for a basic IPv6 ACL

The following shows how your switch performs depth-first match in a basic IPv6 ACL:

- 1 Sort rules by source IPv6 address wildcard first and compare packets against the rule configured with a longer prefix in the source IPv6 address wildcard prior to other rules.
- 2 If two rules are present with the same prefix length in their source IPv6 address wildcards, compare packets against the rule configured first prior to the other.

Depth-first match for an advanced IPv6 ACL

The following shows how your switch performs depth-first match in an advanced IPv6 ACL:

- 1 Sort rules by protocol range first, and compare packets against the rule with the protocol carried on IPv6 specified prior to other rules.
- 2 If two rules are present with the same protocol range, look at source IPv6 address wildcard in addition. Then, compare packets against the rule configured with a larger prefix length in the source IPv6 address wildcard prior to the other.
- 3 If the prefix lengths in the source IPv6 address wildcards are the same, look at the destination IPv6 address wildcards. Then, compare packets against the rule configured with a larger prefix length in the destination IPv6 address wildcard prior to the other.
- 4 If the prefix lengths in the destination IPv6 address wildcards are the same, look at the Layer 4 port number (TCP/UDP port number). Then compare packets against the rule configured with the lower port number prior to the other.
- 5 If the port numbers are the same, compare packets against the rule configured first prior to the other.

The comparison of a packet against an ACL stops once a match is found. The packet is then processed as per the rule.

IPv6 ACL Step Refer to “IPv4 ACL Step” on page 837.

Effective Period of an IPv6 ACL Refer to “Effective Period of an IPv4 ACL” on page 838.

63

IPv4 ACL CONFIGURATION

When configuring an IPv4 ACL, go to these sections for information you are interested in:

- “Creating a Time Range” on page 851
- “Configuring a Basic IPv4 ACL” on page 842
- “Configuring an Advanced IPv4 ACL” on page 844
- “Configuring an Ethernet Frame Header ACL” on page 845
- “Copying an IPv4 ACL” on page 846
- “Displaying and Maintaining IPv4 ACLs” on page 847
- “IPv4 ACL Configuration Example” on page 847

Creating a Time Range

You can specify a time range for each rule in an ACL. A time range-based ACL takes effect only in specified time ranges. Only after a time range is configured and the system time is within the time range, can an ACL rule take effect.

Two types of time ranges are available:

- Periodic time range, which recurs periodically on the day or days of the week.
- Absolute time range, which takes effect only in a period of time and does not recur.

Configuration Procedure

Follow these steps to create a time range:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Create a time range	time-range <i>time-name</i> { <i>start-time to end-time</i> days [from <i>time1 date1</i>] [to <i>time2 date2</i>] from <i>time1 date1</i> [to <i>time2 date2</i>] to <i>time2 date2</i> }	Required



- *Periodic time range created using the **time-range** *time-name* *start-time to end-time* days command. A time range thus created recurs periodically on the day or days of the week.*
- *Absolute time range created using the **time-range** *time-name* { **from** *time1 date1* [**to** *time2 date2*] | **to** *time2 date2* } command. Unlike a periodic time range, a time range thus created does not recur. For example, to create an absolute time range that is active between January 1, 2004 00:00 and December 31, 2004 23:59, you may use the **time-range test from 00:00 01/01/2004 to 23:59 12/31/2004** command.*

- Compound time range created using the **time-range** time-name start-time to end-time days { **from** time1 date1 [**to** time2 date2] | **to** time2 date2 } command. A time range thus created recurs on the day or days of the week only within the specified period. For example, to create a time range that is active from 12:00 to 14:00 on Wednesdays between January 1, 2004 00:00 and December 31, 2004 23:59, you may use the **time-range test 12:00 to 14:00 wednesday from 00:00 01/01/2004 to 23:59 12/31/2004** command.
- You may create individual time ranges identified with the same name. They are regarded as one time range whose active period is the result of ORing periodic ones, ORing absolute ones, and ANDing periodic and absolute ones.
- With no start time specified, the time range is from the earliest time that the system can express (that is, 00:00 01/01/1970) to the end time. With no end time specified, the time range is from the time the configuration takes effect to the latest time that the system can express (that is, 24:00 12/31/2100).
- Up to 256 time ranges can be defined.

Configuration Examples # Create a periodic time range that is active from 8:00 to 18:00 every working day.

```
<Sysname> system-view
[Sysname] time-range test 8:00 to 18:00 working-day
[Sysname] display time-range test
Current time is 22:17:42 1/5/2006 Thursday
```

```
Time-range : test ( Inactive )
08:00 to 18:00 working-day
```

Create an absolute time range from 15:00, Jan 28, 2006 to 15:00, Jan 28, 2008.

```
<Sysname> system-view
[Sysname] time-range test from 15:00 1/28/2006 to 15:00 1/28/2008
[Sysname] display time-range test
Current time is 22:20:18 1/5/2006 Thursday
```

```
Time-range : test ( Inactive )
from 15:00 1/28/2006 to 15:00 1/28/2008
```

Configuring a Basic IPv4 ACL

Basic IPv4 ACLs filter packets based on source IP address. They are numbered in the range 2000 to 2999.

Configuration Prerequisites

If you want to reference a time range to a rule, define it with the **time-range** command first.

Configuration Procedure

Follow these steps to configure a basic IPv4 ACL:

To do...	Use the command...	Remarks
Enter system view	system-view	--

To do...	Use the command...	Remarks
Create and enter basic IPv4 ACL view	acl number <i>acl-number</i> [name <i>acl-name</i>] [match-order { auto config }]	Required The default match order is config . If you specify a name for an IPv4 ACL when creating the ACL, you can use the acl name <i>acl-name</i> command to enter the view of the ACL later.
Create or modify a rule	rule [<i>rule-id</i>] { deny permit } [fragment logging source { <i>sour-addr</i> <i>sour-wildcard</i> any } time-range <i>time-name</i>] *	Required To create multiple rules, repeat this step. Note that the logging keyword is not supported if the ACL is to be referenced by a QoS policy for traffic classification.
Set a rule numbering step	step <i>step-value</i>	Optional The default step is 5.
Create an IPv4 ACL description	description <i>text</i>	Optional By default, no IPv4 ACL description is present.
Create a rule description	rule <i>rule-id</i> comment <i>text</i>	Optional By default, no rule description is present.



- You will fail to create or modify a rule if its permit/deny statement is exactly the same as another rule. In addition, if the ACL match order is set to **auto** rather than **config**, you cannot modify ACL rules.
- You may use the **display acl** command to verify rules configured in an ACL. If the match order for this ACL is **auto**, rules are displayed in the depth-first match order rather than by rule number.



CAUTION:

- You can modify the match order of an ACL with the **acl number** *acl-number* [**name** *acl-name*] **match-order** { **auto** | **config** } command but only when it does not contain any rules.
- The rule specified in the **rule comment** command must have existed.

Configuration Examples

Create IPv4 ACL 2000 to deny the packets with source address 1.1.1.1 to pass.

```
<Sysname> system-view
[Sysname] acl number 2000
[Sysname-acl-basic-2000] rule deny source 1.1.1.1 0
```

Verify the configuration.

```
[Sysname-acl-basic-2000] display acl 2000
Basic ACL 2000, named -none-, 1 rule,
ACL's step is 5
rule 0 deny source 1.1.1.1 0
```

Configuring an Advanced IPv4 ACL

Advanced IPv4 ACLs filter packets based on source IP address, destination IP address, protocol carried on IP, and other protocol header fields, such as the TCP/UDP source port, TCP/UDP destination port, ICMP message type, and ICMP message code.

In addition, advanced IPv4 ACLs allow you to filter packets based on three priority criteria: type of service (ToS), IP precedence, and differentiated services codepoint (DSCP) priority.

Advanced IPv4 ACLs are numbered in the range 3000 to 3999. Compared with basic IPv4 ACLs, they allow of more flexible and accurate filtering.

Configuration Prerequisites

If you want to reference a time range to a rule, define it with the **time-range** command first.

Configuration Procedure

Follow these steps to configure an advanced IPv4 ACL:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Create and enter advanced IPv4 ACL view	acl number <i>acl-number</i> [name <i>acl-name</i>] [match-order { auto config }]	Required The default match order is config . If you specify a name for an IPv4 ACL when creating the ACL, you can use the acl name <i>acl-name</i> command to enter the view of the ACL later.
Create or modify a rule	rule [<i>rule-id</i>] { deny permit } protocol [destination { <i>dest-addr</i> <i>dest-wildcard</i> any } destination-port <i>operator</i> <i>port1</i> [<i>port2</i>] dscp <i>dscp</i> established fragment icmp-type { <i>icmp-type</i> <i>icmp-code</i> <i>icmp-message</i> } logging precedence <i>precedence</i> reflective source { <i>sour-addr</i> <i>sour-wildcard</i> any } source-port <i>operator</i> <i>port1</i> [<i>port2</i>] time-range <i>time-name</i> tos <i>tos</i>] *	Required To create multiple rules, repeat this step. Note that if the ACL is to be referenced by a QoS policy for traffic classification, the logging and reflective keywords are not supported and the <i>operator</i> argument cannot be: <ul style="list-style-type: none"> ■ neq, if the policy is for the inbound traffic, ■ gt, lt, neq or range, if the policy is for the outbound traffic.
Set a rule numbering step	step <i>step-value</i>	Optional The default step is 5.
Create an IPv4 ACL description	description <i>text</i>	Optional By default, no IPv4 ACL description is present.
Create a rule description	rule <i>rule-id</i> comment <i>text</i>	Optional By default, no rule description is present.



- You will fail to create or modify a rule if its permit/deny statement is exactly the same as another rule. In addition, if the ACL match order is set to **auto** rather than **config**, you cannot modify ACL rules.
- You may use the **display acl** command to verify rules configured in an ACL. If the match order for this ACL is **auto**, rules are displayed in the depth-first match order rather than by rule number.

**CAUTION:**

- You can modify the match order of an ACL with the **acl number** *acl-number* [**name** *acl-name*] **match-order** { **auto** | **config** } command but only when it does not contain any rules.
- The rule specified in the **rule comment** command must have existed.

Configuration Examples

Create IPv4 ACL 3000, permitting TCP packets with port number 80 sent from 129.9.0.0 to 202.38.160.0 to pass.

```
<Sysname> system-view
[Sysname] acl number 3000
[Sysname-acl-adv-3000] rule permit tcp source 129.9.0.0 0.0.255.255 destination 202.38.160.0 0.0.0.255 destination-port eq 80
```

Verify the configuration.

```
[Sysname-acl-adv-3000] display acl 3000
Advanced ACL 3000, named -none-, 1 rule,
ACL's step is 5
rule 0 permit tcp source 129.9.0.0 0.0.255.255 destination 202.38.160.0 0.0.0.255 destination-port eq www
```

Configuring an Ethernet Frame Header ACL

Ethernet frame header ACLs filter packets based on Layer 2 protocol header fields such as source MAC address, destination MAC address, 802.1p priority (VLAN priority), and link layer protocol type. They are numbered in the range 4000 to 4999.

Configuration Prerequisites

If you want to reference a time range to a rule, define it with the **time-range** command first.

Configuration Procedure

Follow these steps to configure an Ethernet frame header ACL:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Create and enter Ethernet frame header ACL view	acl number <i>acl-number</i> [name <i>acl-name</i>] [match-order { auto config }]	Required The default match order is config . If you specify a name for an IPv4 ACL when creating the ACL, you can use the acl name <i>acl-name</i> command to enter the view of the ACL later.

To do...	Use the command...	Remarks
Create or modify a rule	rule [<i>rule-id</i>] { deny permit } [cos <i>vlan-pri</i>] dest-mac <i>dest-addr</i> <i>dest-mask</i> lsap <i>lsap-code</i> <i>lsap-wildcard</i> source-mac <i>sour-addr</i> <i>source-mask</i> time-range <i>time-name</i> type <i>type-code</i> <i>type-wildcard</i>] *	Required To create multiple rules, repeat this step. Note that the lsap keyword is not supported if the ACL is to be referenced by a QoS policy for traffic classification.
Set a rule numbering step	step <i>step-value</i>	Optional The default step is 5.
Create an ACL description	description <i>text</i>	Optional By default, no IPv4 ACL description is present.
Create a rule description	rule <i>rule-id</i> comment <i>text</i>	Optional By default, no rule description is present.



- You will fail to create or modify a rule if its *permit/deny* statement is exactly the same as another rule. In addition, if the ACL match order is set to **auto** rather than **config**, you cannot modify ACL rules.
- You may use the **display acl** command to verify rules configured in an ACL. If the match order for this ACL is **auto**, rules are displayed in the depth-first match order rather than by rule number.



CAUTION:

- You can modify the match order of an ACL with the **acl number** *acl-number* [**name** *acl-name*] **match-order** { **auto** | **config** } command but only when it does not contain any rules.
- The rule specified in the **rule comment** command must have existed.

Configuration Examples

Create ACL 4000 to deny frames with the 802.1p priority of 3.

```
<Sysname> system-view
[Sysname] acl number 4000
[Sysname-acl-ethernetframe-4000] rule deny cos 3
```

Verify the configuration.

```
[Sysname-acl-ethernetframe-4000] display acl 4000
Ethernet frame ACL 4000, named -none-, 1 rule,
ACL's step is 5
rule 0 deny cos excellent-effort
```

Copying an IPv4 ACL

This feature allows you to copy an existent IPv4 ACL to generate a new one, which is of the same type and has the same match order, match rules, rule numbering step and descriptions as the source IPv4 ACL.

Configuration Prerequisites

Make sure that the source IPv4 ACL exists while the destination IPv4 ACL does not.

Configuration Procedure Follow these steps to copy an IPv4 ACL:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Copy an existing IPv4 ACL to generate a new one of the same type	acl copy { <i>source-acl-number</i> name <i>source-acl-name</i> } to { <i>dest-acl-number</i> name <i>dest-acl-name</i> }	Required



CAUTION:

- *The source IPv4 ACL and the destination IPv4 ACL must be of the same type.*
- *The generated ACL does not take the name of the source IPv4 ACL.*

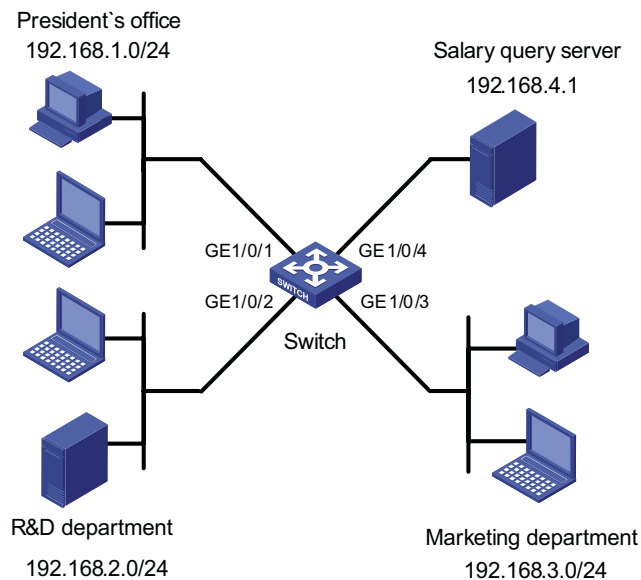
Displaying and Maintaining IPv4 ACLs

To do...	Use the command...	Remarks
Display information about a specified or all IPv4 ACLs	display acl { <i>acl-number</i> all name <i>acl-name</i> }	Available in any view
Display information about ACL uses of a switch	display acl resource	Available in any view
Display the configuration and state of a specified or all time ranges	display time-range { <i>time-name</i> all }	Available in any view
Clear statistics about a specified or all IPv4 ACLs that are referenced by upper layer software	reset acl counter { <i>acl-number</i> all name <i>acl-name</i> }	Available in user view

IPv4 ACL Configuration Example

Network Requirements As shown in Figure 253, a company interconnects its departments through the switch.

Configure an ACL to deny access of all departments but the President's office to the salary query server during office hours (from 8:00 to 18:00) in working days.

Network Diagram **Figure 253** Network diagram for IPv4 ACL configuration

Configuration Procedure

1 Create a time range for office hours

Create a periodic time range spanning 8:00 to 18:00 in working days.

```
<Switch> system-view
[Switch] time-range trname 8:00 to 18:00 working-day
```

2 Define an ACL to control access to the salary query server

Configure a rule to control access of the R&D Department to the salary query server.

```
[Switch] acl number 3000
[Switch-acl-adv-3000] rule deny ip source 192.168.2.0 0.0.0.255 destination 192.168.4.1 0.0.0.0 time-range trname
[Switch-acl-adv-3000] quit
```

Configure a rule to control access of the Marketing Department to the salary query server.

```
[Switch] acl number 3001
[Switch-acl-adv-3001] rule deny ip source 192.168.3.0 0.0.0.255 destination 192.168.4.1 0.0.0.0 time-range trname
[Switch-acl-adv-3001] quit
```

3 Apply the IPv4 ACL

Configure class c_rd for packets matching IPv4 ACL 3000.

```
[Switch] traffic classifier c_rd
[Switch-classifier-c_rd] if-match acl 3000
[Switch-classifier-c_rd] quit
```

Configure traffic behavior b_rd to deny matching packets.

```
[Switch] traffic behavior b_rd
[Switch-behavior-b_rd] filter deny
[Switch-behavior-b_rd] quit
```

Configure class c_market for packets matching IPv4 ACL 3001.

```
[Switch] traffic classifier c_market
[Switch-classifier-c_market] if-match acl 3001
[Switch-classifier-c_market] quit
```

Configure traffic behavior b_market to deny matching packets.

```
[Switch] traffic behavior b_market
[Switch-behavior-b_market] filter deny
[Switch-behavior-b_market] quit
```

Configure QoS policy p_rd to use traffic behavior b_rd for class c_rd.

```
[Switch] qos policy p_rd
[Switch-qospolicy-p_rd] classifier c_rd behavior b_rd
[Switch-qospolicy-p_rd] quit
```

Configure QoS policy p_market to use traffic behavior b_market for class c_market.

```
[Switch] qos policy p_market
[Switch-qospolicy-p_market] classifier c_market behavior b_market
[Switch-qospolicy-p_market] quit
```

Apply QoS policy p_rd to interface GigabitEthernet 1/0/2.

```
[Switch] interface GigabitEthernet 1/0/2
[Switch-GigabitEthernet1/0/2] qos apply policy p_rd inbound
[Switch-GigabitEthernet1/0/2] quit
```

Apply QoS policy p_market to interface GigabitEthernet 1/0/3.

```
[Switch] interface GigabitEthernet 1/0/3
[Switch-GigabitEthernet1/0/3] qos apply policy p_market inbound
```


64

IPv6 ACL CONFIGURATION

When configuring IPv6 ACLs, go to these sections for information you are interested in:

- “Creating a Time Range” on page 851
- “Configuring a Basic IPv6 ACL” on page 851
- “Configuring an Advanced IPv6 ACL” on page 852
- “Copying an IPv6 ACL” on page 854
- “Displaying and Maintaining IPv6 ACLs” on page 854
- “IPv6 ACL Configuration Example” on page 854

Creating a Time Range Refer to section “Creating a Time Range” on page 841.

Configuring a Basic IPv6 ACL Basic IPv6 ACLs filter packets based on source IPv6 address. They are numbered in the range 2000 to 2999.

Configuration Prerequisites If you want to reference a time range to a rule, define it with the **time-range** command first.

Configuration Procedure Follow these steps to configure a basic IPv6 ACL:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Create and enter basic IPv6 ACL view	acl ipv6 number <i>acl6-number</i> [name <i>acl6-name</i>] [match-order { auto config }]	Required The default match order is config . If you specify a name for an IPv6 ACL when creating the ACL, you can use the acl ipv6 name <i>acl6-name</i> command to enter the view of the ACL later.
Create or modify a rule	rule [<i>rule-id</i>] { deny permit } [fragment logging source { <i>ipv6-address</i> <i>prefix-length</i> <i>ipv6-address/prefix-length</i> } any } time-range <i>time-name</i>] *	Required To create multiple rules, repeat this step. Note that the logging and fragment keywords are not supported if the ACL is to be referenced by a QoS policy for traffic classification.
Set a rule numbering step	step <i>step-value</i>	Optional The default step is 5.

To do...	Use the command...	Remarks
Create an IPv6 ACL description	description <i>text</i>	Optional By default, no IPv6 ACL description is present.
Create a rule description	rule <i>rule-id</i> comment <i>text</i>	Optional By default, no rule description is present.



- You will fail to create or modify a rule if its permit/deny statement is exactly the same as another rule. In addition, if the ACL match order is set to **auto** rather than **config**, you cannot modify ACL rules.
- You may use the **display acl** command to verify rules configured in an ACL. If the match order for this ACL is **auto**, rules are displayed in the depth-first match order rather than by rule number.



CAUTION:

- You can modify the match order of an IPv6 ACL with the **acl ipv6 number acl6-number [name acl6-name] match-order { auto | config }** command but only when it does not contain any rules.
- The rule specified in the **rule comment** command must have existed.

Configuration Examples

```
# Create IPv6 ACL 2000 to permit IPv6 packets with source address
2030:5060::9050/64 to pass while denying IPv6 packets with source address
fe80:5060::8050/96.
```

```
<Sysname> system-view
[Sysname] acl ipv6 number 2000
[Sysname-acl6-basic-2000] rule permit source 2030:5060::9050/64
[Sysname-acl6-basic-2000] rule deny source fe80:5060::8050/96
```

```
# Verify the configuration.
```

```
[Sysname-acl6-basic-2000] display acl ipv6 2000
Basic IPv6 ACL 2000, named -none-, 2 rules,
ACL's step is 5
rule 0 permit source 2030:5060::9050/64
rule 5 deny source FE80:5060::8050/96
```

Configuring an Advanced IPv6 ACL

Advanced ACLs filter packets based on the source IPv6 address, destination IPv6 address, protocol carried on IPv6, and other protocol header fields such as the TCP/UDP source port, TCP/UDP destination port, ICMP message type, and ICMP message code.

Advanced IPv6 ACLs are numbered in the range 3000 to 3999. Compared with basic IPv6 ACLs, they allow of more flexible and accurate filtering.

Configuration Prerequisites

If you want to reference a time range to a rule, define it with the **time-range** command first.

Configuration Procedure

Follow these steps to configure an advanced IPv6 ACL:

To do...	Use the command...	Remarks
Enter system view	system-view	--
Create and enter advanced IPv6 ACL view	acl ipv6 number <i>acl6-number</i> [name <i>acl6-name</i>] [match-order { auto config }]	Required The default match order is config . If you specify a name for an IPv6 ACL when creating the ACL, you can use the acl ipv6 name <i>acl6-name</i> command to enter the view of the ACL later.
Create or modify a rule	rule [<i>rule-id</i>] { deny permit } <i>protocol</i> [destination { <i>dest</i> <i>dest-prefix</i> <i>dest/dest-prefix</i> any } destination-port <i>operator port1</i> [<i>port2</i>]] dscp <i>dscp</i> fragment icmpv6-type { <i>icmpv6-type</i> <i>icmpv6-code</i> <i>icmpv6-message</i> } logging source { <i>source source-prefix</i> <i>source/source-prefix</i> any } source-port <i>operator port1</i> [<i>port2</i>]] time-range <i>time-name</i>] *	Required To create multiple rules, repeat this step. Note that if the ACL is to be referenced by a QoS policy for traffic classification, the logging and fragment keywords are not supported and the <i>operator</i> argument cannot be: <ul style="list-style-type: none"> ■ neq, if the policy is for the inbound traffic, ■ gt, lt, neq or range, if the policy is for the outbound traffic.
Set a rule numbering step	step <i>step-value</i>	Optional The default step is 5.
Create an ACL description	description <i>text</i>	Optional By default, no IPv6 ACL description is present.
Create a rule description	rule <i>rule-id</i> comment <i>text</i>	Optional By default, no rule description is present.



- You will fail to create or modify a rule if its permit/deny statement is exactly the same as another rule. In addition, if the ACL match order is set to **auto** rather than **config**, you cannot modify ACL rules.
- You may use the **display acl** command to verify rules configured in an ACL. If the match order for this ACL is **auto**, rules are displayed in the depth-first match order rather than by rule number.



CAUTION:

- You can modify the match order of an IPv6 ACL with the **acl ipv6 number** *acl6-number* [**name** *acl6-name*] **match-order** { **auto** | **config** } command but only when it does not contain any rules.
- The rule specified in the **rule comment** command must have existed.

Configuration Examples

```
# Create IPv6 ACL 3000 to permit the TCP packets with the source address
2030:5060::9050/64 to pass.
```

```
<Sysname> system-view
[Sysname] acl ipv6 number 3000
[Sysname-acl6-adv-3000] rule permit tcp source 2030:5060::9050/64
```

Verify the configuration.

```
[Sysname-acl6-adv-3000] display acl ipv6 3000
Advanced IPv6 ACL 3000, named -none-, 1 rule,
ACL's step is 5
rule 0 permit tcp source 2030:5060::9050/64
```

Copying an IPv6 ACL

This feature allows you to copy an existent IPv6 ACL to generate a new one, which is of the same type and has the same match order, match rules, rule numbering step and descriptions as the source IPv6 ACL.

Configuration Prerequisites

Make sure that the source IPv4 ACL exists while the destination IPv4 ACL does not.

Configuration Procedure

Follow these steps to copy an IPv6 ACL:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Copy an existing IPv6 ACL to generate a new one of the same type	acl ipv6 copy { <i>source-acl6-number</i> name <i>source-acl6-name</i> } to { <i>dest-acl6-number</i> name <i>dest-acl6-name</i> }	Required



CAUTION:

- The source IPv6 ACL and the destination IPv6 ACL must be of the same type.
- The generated IPv6 ACL does not take the name of the source IPv6 ACL.

Displaying and Maintaining IPv6 ACLs

To do...	Use the command...	Remarks
Display information about a specified or all IPv6 ACLs	display acl ipv6 { <i>acl6-number</i> all name <i>acl6-name</i> }	Available in any view
Display information about ACL uses of a switch	display acl resource	Available in any view
Display the configuration and status on time range	display time-range { <i>time-name</i> all }	Available in any view
Clear statistics about a specified or all IPv6 ACLs that are referenced by upper layer software	reset acl ipv6 counter { <i>acl6-number</i> all name <i>acl6-name</i> }	Available in user view

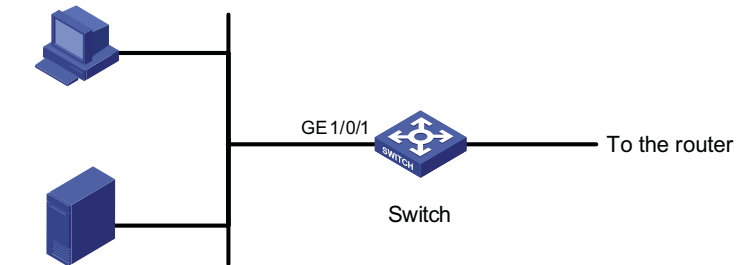
IPv6 ACL Configuration Example

Network Requirements

As shown in Figure 254, a company interconnects its departments through the switch.

Configure an ACL to deny access of the R&D department to external networks.

Network Diagram **Figure 254** Network diagram for IPv6 ACL configuration



R&D department
4050::9000/120

Configuration Procedure # Create an IPv6 ACL 2000.

```
<Switch> system-view
[Switch] acl ipv6 number 2000
[Switch-acl6-basic-2000] rule deny source 4050::9000/120
[Switch-acl6-basic-2000] quit
```

Configure class c_rd for packets matching IPv6 ACL 2000.

```
[Switch] traffic classifier c_rd
[Switch-classifier-c_rd] if-match acl ipv6 2000
[Switch-classifier-c_rd] quit
```

Configure traffic behavior b_rd to deny matching packets.

```
[Switch] traffic behavior b_rd
[Switch-behavior-b_rd] filter deny
[Switch-behavior-b_rd] quit
```

Configure QoS policy p_rd to use traffic behavior b_rd for class c_rd.

```
[Switch] qos policy p_rd
[Switch-qospolicy-p_rd] classifier c_rd behavior b_rd
[Switch-qospolicy-p_rd] quit
```

Apply QoS policy p_rd to interface GigabitEthernet 1/0/1.

```
[Switch] interface GigabitEthernet 1/0/1
[Switch-GigabitEthernet1/0/1] qos apply policy p_rd inbound
```

Introduction

Quality of Service (QoS) is a concept generally existing in occasions where service supply-demand relations exist. QoS measures the ability to meet the service needs of customers. Generally, the evaluation is not to give precise grading. The purpose of the evaluation is to analyze the conditions where the services are good and the conditions where the services still need to be improved, so that specific improvements can be implemented.

In Internet, QoS measures the ability of the network to deliver packets. The evaluation on QoS can be based on different aspects because the network provides diversified services. Generally speaking, QoS is the evaluation on the service ability to support the critical indexes such as delay, delay jitter and packet loss rate in packet delivery.

Traditional Packet Forwarding Service

In traditional IP networks, packets are treated equally. That is, the FIFO (first in first out) policy is adopted for packet processing. Network resources required for packet forwarding is determined by the order in which packets arrive. All the packets share the resources of the network. Network resources available to the packets completely depend on the time they arrive. This service policy is known as Best-effort, which delivers the packets to their destination with the best effort, with no assurance and guarantee for delivery delay, jitter, packet loss ratio, reliability, and so on.

The traditional Best-Effort service policy is only suitable for applications insensitive to bandwidth and delay, such as WWW, FTP and E-mail.

New Requirements Brought forth by New Services

With the fast development of computer networks, more and more networks are connected into Internet. Internet extends very quickly in scale, coverage and the number of users. More and more users use the Internet as a platform for data transmission and develop various applications on it.

Besides traditional applications such as WWW, FTP, and E-mail, Internet users also try to develop new services on Internet, such as tele-education, tele-medicine, video phones, video conferencing, and video on demand (VOD). Enterprise users also hope to connect their branch offices in different locations through the VPN technology to develop some transaction applications, such as to access to the database of the company or to manage remote switches through Telnet.

The new services have one thing in common: they all have special requirements for delivery performances such as bandwidth, delay, and delay jitter. For example, video conferencing and VOD require the guarantee of high bandwidth, low delay and low delay jitter. Some key services such as the transaction handling and the

Telnet do not necessarily require high bandwidth but they are highly dependent on low delay and need to be processed preferentially in case of congestion.

The emergence of new services brings forward higher requirements for the service capability of the IP network. In the delivery process, users hope to get better services, such as dedicated bandwidth for users, reduced packet loss rate, management and avoidance of network congestion, control of network traffic, provision of packet priority, and so on, instead of just having packets delivered to the destination. To meet these requirements, the network service capability need to be further improved.

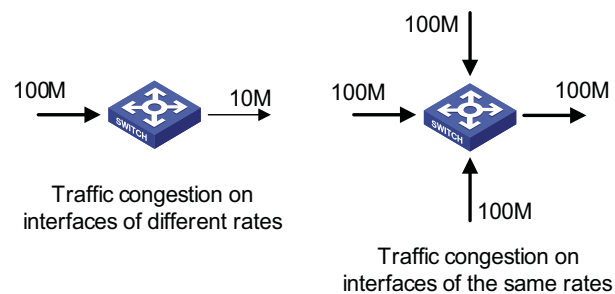
Occurrence and Influence of Congestion and the Countermeasures

QoS issues that traditional networks face are mainly caused by congestion. Congestion means reduced service rate and extra delay introduced because of relatively insufficient resource provisioned.

Occurrence of Congestion

Congestion is very common in a complicated environment of packet switching on Internet. The diagram below gives two examples:

Figure 255 Traffic congestion



- 1 Packets enter a switch over a high-speed link and are forwarded out over a low-speed link.
- 2 Packets enter a switch through multiple interfaces of the same rate at the same time and are forwarded out on an interface of the same rate.

If the outbound traffic exceeds the line rate, the traffic encounters the bottleneck of resources and congestion occurs.

Besides bandwidth bottleneck, any insufficiency of resources for packet forwarding, such as insufficiency of assignable processor time, buffer size, and memory resources can cause congestion. In addition, congestion will also occur if the traffic that arrives within a certain period of time is improperly controlled and the traffic goes beyond the assignable network resources.

Influence of Congestion

Congestion may cause a series of negative influences:

- Congestion increases delay and delay jitter in packet delivery.
- Excessively high delay will cause retransmission of packets.
- Congestion decreases the effective throughput of the network and the utilization of the network resources.

- Aggravated congestion will consume a large amount of network resources (especially memory resources), and unreasonable resource assignment will even lead to system resource deadlock and cause the system breakdown.

It is obvious that congestion is the root of service performance declination because congestion makes traffic unable to get resources timely. However, congestion is common in a complicated environment where packet switching and multi-user services coexist. Therefore, congestion must be treated carefully.

Countermeasures Increasing network bandwidth is a direct way to solve the problem of resource insufficiency, but it cannot solve all the problems that cause network congestion.

A more effective way to solve network congestion problems is to enhance the function of the network layer in traffic control and resource assignment, to provide differentiated services for different requirements, and to assign and utilize resources correctly. In the process of resource assignment and traffic control, the direct or indirect factors that may cause network congestion must be properly controlled so as to reduce the probability of congestion. When congestion occurs, the resource assignment should be balanced according to the features and requirements of all the services to minimize the influence of congestion on QoS.

Major Traffic Management Techniques

Traffic classification, traffic policing (TP), traffic shaping (TS), congestion management, and congestion avoidance are the foundation for providing differentiated services. Their main functions are as follows:

- Traffic classification: Identifies packets according to certain match rules. Traffic classification is the prerequisite of providing differentiated services.
- TP: Monitors and controls the specifications of specific traffic entering the device. When the traffic exceeds the threshold, restrictive or punitive measures can be taken to protect the business interests and network resources of the operator from being damaged.
- Congestion management: Congestion management is necessary for solving resource competition. Congestion management is generally to cache packets in the queues and arrange the forwarding sequence of the packets based on a certain scheduling algorithm.
- Congestion avoidance: Excessive congestion will impair the network resources. Congestion avoidance is to supervise the network resource usage. When it is found that congestion is likely to become worse, the congestion avoidance mechanism will drop packets and regulate traffic to solve the overload of the network.
- TS: TS is a traffic control measure to regulate the output rate of the traffic actively. TS regulates the traffic to match the network resources that can be provided by the downstream devices so as to avoid unnecessary packet loss and congestion.

Among the traffic management techniques, traffic classification is the basis because it identifies packets according to certain match rules, which is the prerequisite of providing differentiated services. TP, TS, congestion management, and congestion avoidance control network traffic and assigned resources from different approaches, and are the concrete ways of providing differentiated services.

66

TRAFFIC CLASSIFICATION, TP, AND LR CONFIGURATION

When configuring traffic classification, TP, and LR, go to these section for information you are interested in:

- "Traffic Classification Overview" on page 861
- "TP and LR Overview" on page 864
- "Traffic Evaluation and Token Bucket" on page 864
- "LR Configuration" on page 866
- "Displaying and Maintaining LR" on page 867

Traffic Classification Overview

Traffic Classification Traffic classification is to identify packets conforming to certain characters according to certain rules. It is the basis and prerequisite for providing differentiated services.

A traffic classification rule can use the precedence bits in the type of service (ToS) field of the IP packet header to identify traffic with different precedence characteristics. A traffic classification rule can also classify traffic according to the traffic classification policy set by the network administrator, such as the combination of source addresses, destination addresses, MAC addresses, IP protocol or the port numbers of the applications. Traffic classification is generally based on the information in the packet header and rarely based on the content of the packet. The classification result is unlimited in range. They can be a small range specified by a quintuplet (source address, source port number, protocol number, destination address, and destination port number), or all the packets to a certain network segment.

Generally, the precedence of bits in the ToS field of the packet header is set when packets are classified on the network border. Thus, IP precedence can be used directly as the classification criterion inside the network. Queue techniques can also process packets differently according to IP precedence. The downstream network can either accept the classification results of the upstream network or re-classify the packets according to its own criterion.

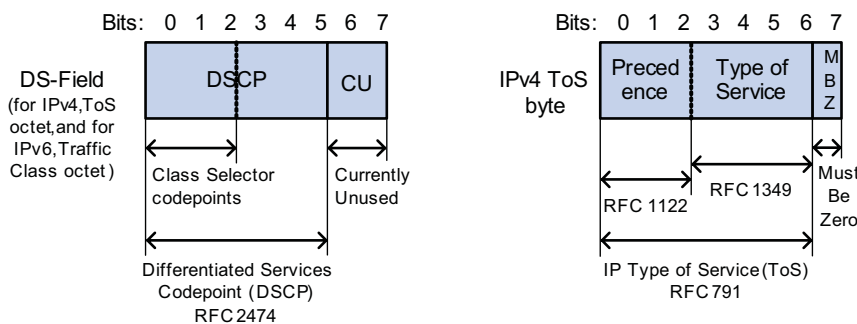
The purpose of traffic classification is to provide differentiated services, so traffic classification is significant only when it is associated with a certain traffic control or resource assignment action. The specific traffic control action to be adopted depends on the phase and the current load status. For example, when the packets enter the network, TP is performed on the packets according to CIR; before the packets flow out of the node, TS is performed on the packets; when congestion

occurs, queue scheduling is performed on the packets; when congestion get worse, congestion avoidance is performed on the packets.

Priority The following describes several types of precedence:

- 1 IP precedence, ToS precedence, and DSCP precedence

Figure 256 DS field and ToS field



The ToS field in an IP header contains eight bits, which are described as follows:

- The first three bits indicate IP precedence in the range of 0 to 7.
- Bit 3 to bit 6 indicate ToS precedence in the range of 0 to 15.
- RFC2474 re-defines the ToS field in the IP packet header, which is called the DS field. The first six (bit 0 to bit 5) bits of the DS field indicate DSCP precedence in the range of 0 to 63. The last two bits (bit 6 and bit 7) are reserved bits.

Table 65 Description on IP Precedence

IP Precedence (decimal)	IP Precedence (binary)	Description
0	000	Routine
1	001	priority
2	010	immediate
3	011	flash
4	100	flash-override
5	101	critical
6	110	internet
7	111	network

In a network providing differentiated services, traffics are grouped into the following four classes, and packets are processed according to their DSCP values.

- Expedited Forwarding (EF) class: In this class, packets can be forwarded regardless of link share of other traffic. The class is suitable for preferential services with low delay, low packet loss ratio, low jitter, and assured bandwidth (such as virtual leased line);
- Assured forwarding (AF) class: This class is further divided into four subclasses (AF1/2/3/4) and a subclass is further divided into three drop priorities, so the AF service level can be segmented. The QoS rank of the AF class is lower than that of the EF class;

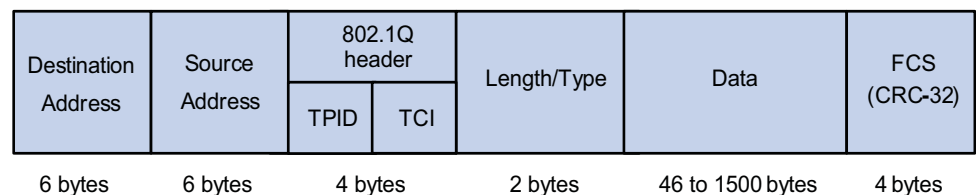
- Class selector (CS) class: This class comes from the IP ToS field and includes eight subclasses;
- Best Effort (BE) class: This class is a special class without any assurance in the CS class. The AF class can be degraded to the BE class if it exceeds the limit. Current IP network traffic belongs to this class by default.

Table 66 Description on DSCP precedence values

DSCP value (decimal)	DSCP value (binary)	Description
46	101110	ef
10	001010	af11
12	001100	af12
14	001110	af13
18	010010	af21
20	010100	af22
22	010110	af23
26	011010	af31
28	011100	af32
30	011110	af33
34	100010	af41
36	100100	af42
38	100110	af43
8	001000	cs1
16	010000	cs2
24	011000	cs3
32	100000	cs4
40	101000	cs5
48	110000	cs6
56	111000	cs7
0	000000	be (default)

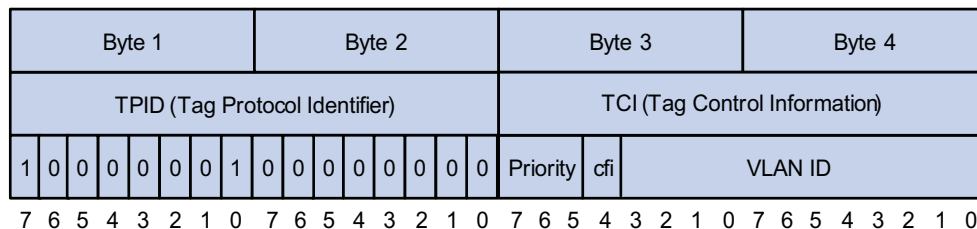
2 802.1p precedence

802.1p precedence lies in Layer 2 packet headers and is applicable to occasions where the Layer 3 packet header does not need analysis but QoS must be assured at Layer 2.

Figure 257 An Ethernet frame with an 802.1Q tag header

As shown in the figure above, the 4-byte 802.1Q tag header contains a 2-byte Tag Protocol Identifier (TPID) whose value is 8100 and a 2-byte Tag Control Information (TCI). TPID is a new class defined by IEEE to indicate a packet with an 802.1Q tag. Figure 258 describes the detailed contents of an 802.1Q tag header.

Figure 258 802.1Q tag headers



In the figure above, the 3-bit priority field in TCI is 802.1p precedence in the range of 0 to 7. In the figure above, the priority field (three bits in length) in TCI is 802.1p precedence (also known as CoS precedence), which ranges from 0 to 7.

Table 67 Description on 802.1p precedence

802.1p precedence (decimal)	802.1p precedence (binary)	Description
0	000	best-effort
1	001	background
2	010	spare
3	011	excellent-effort
4	100	controlled-load
5	101	video
6	110	voice
7	111	network-management

The precedence is called 802.1p precedence because the related applications of this precedence are defined in detail in the 802.1p specifications.

TP and LR Overview

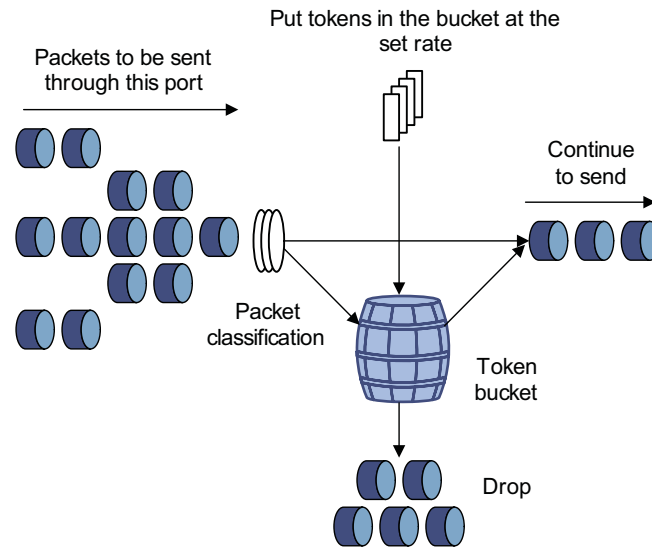
If the traffic from users is not limited, a large amount of continuous burst packets will result in worse network congestion. The traffic of users must be limited in order to make better use of the limited network resources and provide better service for more users. For example, if a traffic flow obtains only the resources committed to it within a certain period of time, network congestion due to excessive burst traffic can be avoided.

TP is traffic control policies for limiting traffic and resource usage by supervising the traffic. The prerequisite for TP is to determine whether or not the traffic exceeds the set threshold. Traffic control policies are adopted only when the traffic exceeds the set threshold. Generally, token bucket is used for evaluating traffic.

Traffic Evaluation and Token Bucket

Token Bucket

A token bucket can be considered as a container with a certain capacity to hold tokens. The system puts tokens into the bucket at a pre-set rate. When the token bucket is full, the extra tokens will overflow and the number of tokens in the bucket stops increasing.

Figure 259 Evaluate traffic with a token bucket

Evaluating Traffic with a Token Bucket

The evaluation for the traffic specification is based on whether the number of tokens in the bucket can meet the need of packet forwarding. If the number of tokens in the bucket is enough to forward the packets, the traffic is conforming to the specification; otherwise, the traffic is nonconforming or excess.

When the token bucket evaluates the traffic, its parameter configurations include:

- Average rate: The rate at which tokens are put into the bucket, namely, the permitted average rate of the traffic. It is generally set to committed information rate (CIR).
- Burst size: The capacity of the token bucket, namely, the maximum traffic size that is permitted in each burst. It is generally set to committed burst size (CBS). The set burst size must be greater than the maximum packet length.

An evaluation is performed on the arrival of each packet. In each evaluation, if the bucket has enough tokens for use, the traffic is controlled within the specification and a number of tokens equivalent to the packet forwarding authority must be taken out; otherwise, this means too many tokens have been used - the traffic is in excess of the specification.

Complicated Evaluation

You can set two token buckets in order to evaluate more complicated conditions and implement more flexible regulation policies. For example, TP uses four parameters:

- CIR
- CBS
- Peak information rate (PIR)
- Excess burst size (EBS)

Two token buckets are used in this evaluation. Their rates of putting tokens into the buckets are CIR and PIR respectively, and their sizes are CBS and EBS respectively (the two buckets are called C bucket and E bucket respectively for short), representing different permitted burst levels. In each evaluation, you can

implement different regulation policies in different conditions, including “enough tokens in C bucket”, “insufficient tokens in C bucket but enough tokens in E bucket” and “insufficient tokens in both C bucket and E bucket”.

TP The typical application of TP is to supervise the specification of certain traffic into the network and limit it within a reasonable range, or to “discipline” the extra traffic. In this way, the network resources and the interests of the operators are protected. For example, you can limit HTTP packets to be within 50% of the network bandwidth. If the traffic of a certain connection is excess, TP can choose to drop the packets or to reset the priority of the packets.

TP is widely used in policing the traffic into the network of internet service providers (ISPs). TP can classify the policed traffic and perform pre-defined policing actions based on different evaluation results. These actions include:

- Forwarding conforming packets or non-conforming packets.
- Dropping conforming or non-conforming packets.
- Marking a conforming packet with a new 802.1p precedence value and forwarding the packet.
- Marking a conforming packet with a new IP precedence value and forwarding the packet.
- Marking a conforming packet or a non-conforming packet with a new DSCP precedence value and forwarding the packet.

LR Port rate limiting refers to limiting the total rate of inbound or outbound packets on a port.

Port rate limiting can be implemented through token buckets. That is, if you perform port rate limiting configuration for a port, the token bucket determines the way to process the packets to be sent by this port or packets reaching the port. Packets can be sent or received if there are enough tokens in the token bucket; otherwise, they will be dropped.

Compared to TP, port rate limiting applies to all the packets passing a port. It is a simpler solution if you want to limit the rate of all the packets passing a port.

LR Configuration

LR Configuration Procedure

Follow these steps to configure LR:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view or port group view	interface <i>interface-type</i> <i>interface-number</i>	Enter either view.
Enter port group view	port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	For Ethernet interface view, the following configuration takes effect only on the current interface. For entering port group view, the following configuration takes effect on all the ports.

To do...	Use the command...	Remarks
Configure LR	qos lr outbound cir <i>committed-information-rate</i> [cbs <i>committed-burst-size</i>]	Required

LR Configuration Examples

Limit the outbound rate of GigabitEthernet 1/0/1 to 640 kbps.

Enter system view

```
<Sysname> system-view
```

Enter interface view

```
[Sysname] interface GigabitEthernet 1/0/1
```

Configure LR parameter and limit the outbound rate to 640 kbps

```
[Sysname-GigabitEthernet1/0/1] qos lr outbound cir 640
```

Displaying and Maintaining LR

To do...	Use the command...	Remarks
Display the LR configuration of an interface	display qos lr interface [<i>interface-type</i> <i>interface-number</i>]	Available in any view

67

QoS POLICY CONFIGURATION

When configuring QoS policy, go to these sections for information that you are interested in:

- "Overview" on page 869
- "Configuring QoS Policy" on page 870
- "Introduction to QoS Policies" on page 870
- "Configuring a QoS Policy" on page 870
- "Displaying and Maintaining QoS Policy" on page 876

Overview

QoS policy includes the following three elements: class, traffic behavior and policy. You can bind the specified class to the specified traffic behavior through QoS policies to facilitate the QoS configuration.

Class

Class is used for identifying traffic.

The elements of a class include the class name and classification rules.

You can use commands to define a series of rules to classify packets. Additionally, you can use commands to define the relationship among classification rules: "**and**" and "**or**".

- **and**: The device considers a packet to be of a specific class when the packet matches all the specified classification rules.
- **or**: The device considers a packet be of a specific class when the packet matches one of the specified classification rules.

Traffic behavior

Traffic behavior is used to define all the QoS actions performed on packets.

The elements of a QoS behavior include traffic behavior name and actions defined in traffic behavior.

You can use commands to define multiple actions in a traffic behavior.

Policy

Policy is used to bind the specified class to the specified traffic behavior.

The elements of a policy include the policy name and the name of the classification-to-behavior binding.

Configuring QoS Policy

The procedure for configuring QoS policy is as follows:

- 1 Define a class and define a group of traffic classification rules in class view.
- 2 Define a traffic behavior and define a group of QoS actions in traffic behavior view.
- 3 Define a policy and specify a traffic behavior corresponding to the class in policy view.
- 4 Apply the QoS policy in Ethernet port view/port group view.

Introduction to QoS Policies

Table 68 QoS policies

Policy name	Corresponding class	Related command
Accounting	Use the if-match <i>match-criteria</i> command to define the class as required for the policy to be associated with.	accounting
TP	Use the if-match <i>match-criteria</i> command to define the class as required for the policy to be associated with.	car
Traffic filtering	Use the if-match <i>match-criteria</i> command to define the class as required for the policy to be associated with.	filter
Traffic mirroring	Use the if-match <i>match-criteria</i> command to define the class as required for the policy to be associated with.	mirror-to
Nested VLAN tag	Use the if-match <i>match-criteria</i> command to define the class as required for the policy to be associated with.	nest
Traffic redirect	Use the if-match <i>match-criteria</i> command to define the class as required for the policy to be associated with.	redirect
Priority marking	Use the if-match <i>match-criteria</i> command to define the class as required for the policy to be associated with.	remark

Configuring a QoS Policy

Configuration Prerequisites

- The name and the rules of the class to which the policy is to be bound to are determined.
- The traffic behavior name and actions in the traffic behavior in the policy are determined.
- The policy name is determined.
- Apply the QoS policy in Ethernet port view/port group view.

Defining a Class

To define a class, you need to create a class and then define rules in the corresponding class view.

Configuration procedure

Follow these steps to define a class:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a class and enter the corresponding class view	traffic classifier <i>classifier-name</i> [operator { and or }]	Required By default, the and keyword is specified. That is, the relation between the rules in the class view is logic AND. This operation leads you to class view.
Define a rule used to match packets	if-match <i>match-criteria</i>	Required

match-criteria: Matching rules to be defined for a class. Table 69 describes the available forms of this argument.

Table 69 The form of the match-criteria argument

Form	Description
acl <i>access-list-number</i>	Specifies an ACL to match packets. The <i>access-list-number</i> argument is in the range 2000 to 4999. In a class configured with the operator and , the logical relationship between rules defined in the referenced IPv4 ACL is or .
acl ipv6 <i>access-list-number</i>	Specifies an IPv6 ACL to match IPv6 packets. The <i>access-list-number</i> argument is in the range 2000 to 3999. In a class configured with the operator and , the logical relationship between rules defined in the referenced IPv6 ACL is or .
any	Specifies to match all packets.
customer-dot1p <i>8021p-list</i>	Specifies to match packets by 802.1p precedence of the customer network. The <i>8021p-list</i> argument is a list of CoS values. You can provide up to eight space-separated CoS values for this argument. CoS is in the range 0 to 7.
customer-vlan-id <i>vlan-id-list</i>	Specifies to match the packets of specified VLANs of user networks. The <i>vlan-id-list</i> argument specifies a list of VLAN IDs, in the form of <i>vlan-id to vlan-id</i> or multiple discontinuous VLAN IDs (separated by space). You can specify up to eight VLAN IDs for this argument at a time. VLAN ID is in the range 1 to 4094. In a class configured with the operator and , the logical relationship between the customer VLAN IDs specified for the customer-vlan-id keyword is or .
destination-mac <i>mac-address</i>	Specifies to match the packets with a specified destination MAC address.
dscp <i>dscp-list</i>	Specifies to match packets by DSCP precedence. The <i>dscp-list</i> argument is a list of DSCP values. You can provide up to eight space-separated DSCP values for this argument. DSCP is in the range of 0 to 63.
ip-precedence <i>ip-precedence-list</i>	Specifies to match packets by IP precedence. The <i>ip-precedence-list</i> argument is a list of IP precedence values. You can provide up to eight space-separated IP precedence values for this argument. IP precedence is in the range 0 to 7.
protocol <i>protocol-name</i>	Specifies to match the packets of a specified protocol. The protocol-name argument can be IP or IPv6.

Table 69 The form of the match-criteria argument

Form	Description
service-dot1p <i>8021p-list</i>	Specifies to match packets by 802.1p precedence of the service provider network. The <i>8021p-list</i> argument is a list of CoS values. You can provide up to eight space-separated CoS values for this argument. CoS is in the range 0 to 7. In a class configured with the operator and , the logical relationship between the service VLAN IDs specified for the service-vlan-id keyword is or .
service-vlan-id <i>vlan-id-list</i>	Specifies to match the packets of the VLANs of the operator's network. The <i>vlan-id-list</i> argument is a list of VLAN IDs, in the form of <i>vlan-id to vlan-id</i> or multiple discontinuous VLAN IDs (separated by space). You can specify up to eight VLAN IDs for this argument at a time. VLAN ID is in the range of 1 to 4094.
source-mac <i>mac-address</i>	Specifies to match the packets with a specified source MAC address.



Suppose the logical relationship between classification rules is **and**. Note the following when using the **if-match** command to define matching rules.

- If multiple matching rules with the **acl** or **acl ipv6** keyword specified are defined in a class, the actual logical relationship between these rules is **or** when the policy is applied.
- If multiple matching rules with the **customer-vlan-id** or **service-vlan-id** keyword specified are defined in a class, the actual logical relationship between these rules is **or**.

Configuration example

1 Network requirements

Configure a class named test to match the packets with their IP precedence being 6.

2 Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Create the class. (This operation leads you to class view.)

```
[Sysname] traffic classifier test
```

Define the classification rule.

```
[Sysname-classifier-test] if-match ip-precedence 6
```

Defining a Traffic Behavior

To define a traffic behavior, you need to create a traffic behavior and then configure attributes for it in traffic behavior view.

Configuration procedure

Follow these steps to define a traffic behavior:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a traffic behavior and enter the corresponding traffic behavior view	traffic behavior <i>behavior-name</i>	Required <i>behavior-name</i> : Behavior name. This operation leads you to traffic behavior view
Configure accounting action	accounting	Required
Configure TP action	car cir <i>committed-information-rate</i> [cbs <i>committed-burst-size</i>] [ebs <i>excess-burst-size</i>] [pir <i>peak-information-rate</i>] [green <i>action</i>] [red <i>action</i>] [yellow <i>action</i>]	You can configure the traffic behavior as required.
Configure traffic filtering behavior	filter { deny permit }	
Configure traffic mirroring action	mirror-to { cpu interface <i>interface-type interface-number</i> }	
Configure nested VLAN tag action	nest top-most vlan-id <i>vlan-id</i>	
Configure traffic redirect action	redirect { cpu interface <i>interface-type interface-number</i> link-aggregation group <i>agg-id</i> next-hop { <i>ipv4-add</i> [<i>ipv4-add</i>] <i>ipv6-add</i> [<i>interface-type interface-number</i>] [<i>ipv6-add</i> [<i>interface-type interface-number</i>]] } }	
Remark the customer network VLAN ID for packets	remark customer-vlan-id <i>vlan-id-value</i>	
Remark DSCP value for packets	remark dscp <i>dscp-value</i>	
Remark 802.1p precedence for packets	remark dot1p <i>8021p</i>	
Remark drop precedence for packets	remark drop-precedence <i>drop-precedence-value</i>	
Remark IP precedence for packets	remark ip-precedence <i>ip-precedence-value</i>	
Remark local precedence for packets	remark local-precedence <i>local-precedence</i>	
Remark the service provider network VLAN ID for packets	remark service-vlan-id <i>vlan-id-value</i>	

Configuration example

1 Network requirements

Create a traffic behavior named test, configuring TP action for it, with the CAR being 640 kbps.

2 Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Create the traffic behavior (This operation leads you to traffic behavior view).

```
[Sysname] traffic behavior test
```

Configure TP action for the traffic behavior.

```
[Sysname-behavior-test] car cir 640
```

Defining a Policy

A policy associates a class with a traffic behavior. Each traffic behavior is comprised of a group of QoS actions. A device executes these QoS actions in the order they are defined.

Follow these steps to associate a traffic behavior with a class:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a policy (This operation leads you to policy view)	qos policy <i>policy-name</i>	-
Specify the traffic behavior for a class	classifier <i>classifier-name</i> behavior <i>behavior-name</i>	Required



In a QoS policy with multiple class-to-traffic-behavior associations, if the action of creating an outer VLAN tag, the action of setting customer network VLAN ID, or the action of setting service provider network VLAN ID is configured in a traffic behavior, we recommend you not to configure any other action in this traffic behavior. Otherwise, the QoS policy may not function as expected after it is applied.

Applying a Policy

Configuration procedure

Follow these steps to apply a policy on a port:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter port view or port group view	interface <i>interface-type</i> <i>interface-number</i> port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Perform either of the two operations. The configuration performed in Ethernet port view applies to the current port only. The configuration performed in port group view applies to all the ports in the port group.
Apply an associated policy	qos apply policy <i>policy-name</i> { inbound outbound }	Required

Note that, when you apply a policy by using the **qos apply policy** command, whether or not the **inbound/outbound** keyword can take effect depends on the actions defined in the traffic behavior, as described in Table 70.

Table 70 The support for the inbound direction and the outbound direction

Action	Inbound	Outbound
Traffic accounting	Supported	Supported

Table 70 The support for the inbound direction and the outbound direction

Action	Inbound	Outbound
TP	Supported	Supported
Traffic filtering	Supported	Supported
Traffic mirroring	Supported	Supported
Configuring the outer VLAN tag	Supported	Not supported
Traffic redirecting	Supported	Not supported
Remarking the customer network VLAN ID for packets	Not supported	Supported
Remarking the 802.1p precedence for packets	Supported	Supported
Remarking the drop precedence for packets	Supported	Not supported
Remarking the DSCP precedence for packets	Supported	Supported
Remarking the IP precedence for packets	Supported	Supported
Remarking the local precedence for packets	Supported	Not supported
Remarking the service provider network VLAN ID for packets	Supported	Supported



CAUTION: Follow these rules when configuring a behavior. Otherwise the corresponding QoS policy cannot be applied successfully.

- The action of creating an outer VLAN tag cannot be configured simultaneously with any other action except the traffic filtering action or the action of setting 802.1p precedence in the same traffic behavior. The action of creating an outer VLAN tag must be applied to basic QinQ-enabled ports or port groups.
- When the action of setting the service provider network VLAN ID is applied in the inbound direction, any other action except the traffic filtering action or the action of setting 802.1p precedence cannot be configured in the same traffic behavior.
- When the action of mirroring traffic is applied in the outbound direction, any other action cannot be configured in the same traffic behavior.

Configuration example

1 Network requirements

Configure a policy named test to associate the traffic behavior named test_behavior with the class named test_class. Apply the policy to the inbound direction of GigabitEthernet 1/0/1 port.

2 Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Create a policy (This operation leads you to policy view).

```
[Sysname] qos policy test
[Sysname-qospolicy-test]
```

Associate the traffic behavior named test_behavior with the class named test_class.

```
[Sysname-qospolicy-test] classifier test_class behavior test_behavior
[Sysname-qospolicy-test] quit
```

Enter port view.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1]
```

Apply the policy to the port.

```
[Sysname-GigabitEthernet1/0/1] qos apply policy test inbound
```

Displaying and Maintaining QoS Policy

To do...	Use the command...	Remarks
Display the information about a class and the corresponding actions associated by a policy	display qos policy user-defined [<i>policy-name</i> [classifier <i>classifier-name</i>]]	Available in any view
Display the information about the policies applied on a port	display qos policy interface [<i>interface-type interface-number</i>] [inbound outbound]	
Display the information about a traffic behavior	display traffic behavior user-defined [<i>behavior-name</i>]	
Display the information about a class	display traffic classifier user-defined [<i>classifier-name</i>]	

When configuring congestion management, go to these section for information that you are interested in:

- "Overview" on page 877
- "Congestion Management Policy" on page 877
- "Configuring an SP Queue" on page 879
- "Configuring a WRR Queue" on page 880
- "Configuring SP+WRR Queues" on page 881
- "Displaying and Maintaining Congestion Management" on page 882

Overview

When the rate at which the packets arrive is higher than the rate at which the packets are transmitted on an interface, congestion occurs on this interface. If there is not enough storage space to store these packets, parts of them will be lost. Packet loss may cause the transmitting device to retransmit the packets because the lost packets time out, which causes a malicious cycle.

The core of congestion management is how to schedule the resources and determine the sequence of forwarding packets when congestion occurs.

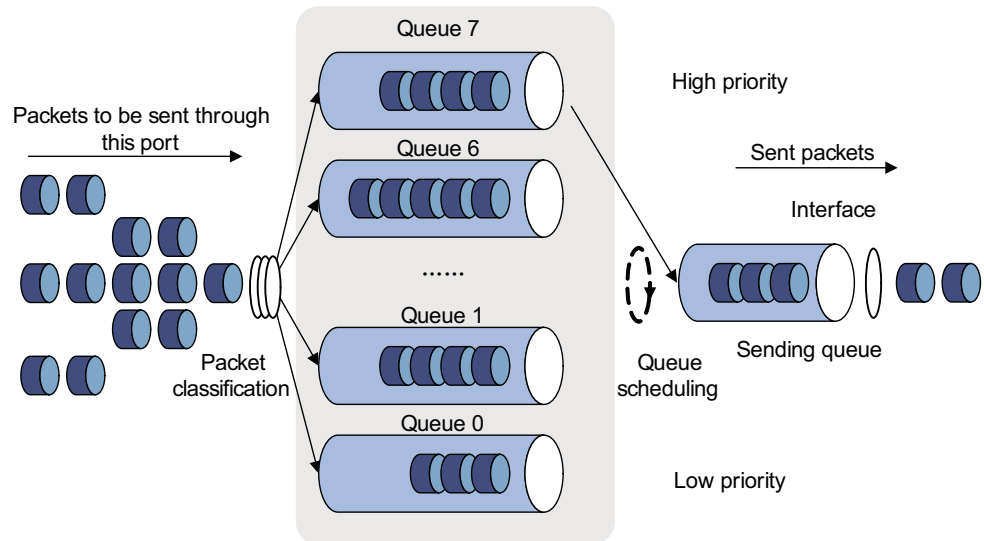
Congestion Management Policy

Queuing technology is generally adopted to solve the congestion problem. The queuing technology is to classify the traffic according to a specified queue-scheduling algorithm and then use the specified priority algorithm to forward the traffic. Each queuing algorithm is used to solve specific network traffic problems and affects the parameters such as bandwidth allocation, delay and delay jitter.

The following paragraphs describe strict-priority (SP) queue-scheduling algorithm, and weighted round robin (WRR) queue-scheduling algorithm.

1 SP queue-scheduling algorithm

Figure 260 Diagram for SP queuing

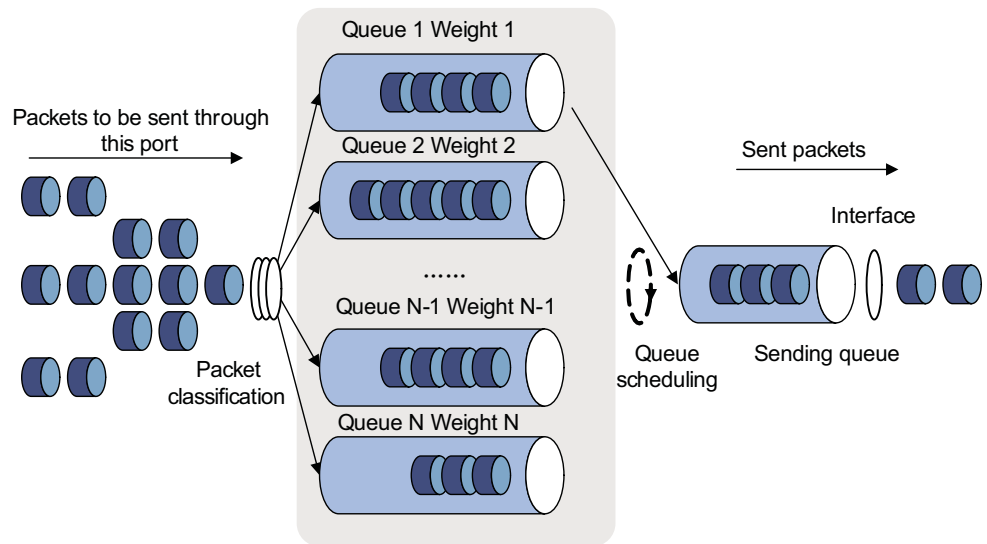


SP queue-scheduling algorithm is specially designed for critical service applications. An important feature of critical services is that they demand preferential service in congestion in order to reduce the response delay. Assume that there are eight output queues on the port and the preferential queue classifies the eight output queues on the port into eight classes, which are queue7, queue6, queue5, queue4, queue3, queue2, queue1, and queue0. Their priorities decrease in order.

In queue scheduling, SP sends packets in the queue with higher priority strictly following the priority order from high to low. When the queue with higher priority is empty, packets in the queue with lower priority are sent. You can put critical service packets into the queues with higher priority and put non-critical service (such as e-mail) packets into the queues with lower priority. In this case, critical service packets are sent preferentially and non-critical service packets are sent when critical service groups are not sent.

The disadvantage of SP queue is that: if there are packets in the queues with higher priority for a long time in congestion, the packets in the queues with lower priority will be "starved" because they are not served.

2 WRR queue-scheduling algorithm

Figure 261 Diagram for WRR queuing

A port of the switch supports eight outbound queues. The WRR queue-scheduling algorithm schedules all the queues in turn to ensure that every queue can be assigned a certain service time. Assume there are eight output queues on the port. The eight weight values (namely, w_7 , w_6 , w_5 , w_4 , w_3 , w_2 , w_1 , and w_0) indicating the proportion of assigned resources are assigned to the eight queues respectively. On a 100M port, you can configure the weight values of WRR queue-scheduling algorithm to 50, 30, 10, 10, 50, 30, 10, and 10 (corresponding to w_7 , w_6 , w_5 , w_4 , w_3 , w_2 , w_1 , and w_0 respectively). In this way, the queue with the lowest priority can be assured of 5 Mbps of bandwidth at least, thus avoiding the disadvantage of SP queue-scheduling algorithm that packets in low-priority queues are possibly not to be served for a long time. Another advantage of WRR queue-scheduling algorithm is that though the queues are scheduled in turn, the service time for each queue is not fixed, that is to say, if a queue is empty, the next queue will be scheduled immediately. In this way, the bandwidth resources are fully utilized.

3Com Switch 4800G Family support the following three queue scheduling algorithms:

- All the queues are scheduled through the SP algorithm.
- All the queues are scheduled through the WRR algorithm.
- Some queues are scheduled through the SP algorithm, while other queues are scheduled through the WRR algorithm.

Configuring an SP Queue

Configuration Procedure Follow these steps to configure SP queues:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks	
Enter port view or port group view	Enter port view Enter port group view	interface <i>interface-type interface-number</i> port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Perform either of the two operations. The configuration performed in Ethernet port view applies to the current port only. The configuration performed in port group view applies to all the ports in the port group.
Configure SP queue scheduling algorithm	qos sp	Required By default, all the ports adopt the WRR queue scheduling algorithm, with the weight values assigned to queue 0 through queue 7 being 1, 2, 3, 4, 5, 9, 13, and 15.	

Configuration Examples Network requirements

Configure GigabitEthernet1/0/1 to adopt SP queue scheduling algorithm.

Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Configure an SP queue for GigabitEthernet1/0/1 port.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] qos sp
```

Configuring a WRR Queue

By default, SP queue scheduling algorithm is adopted on all the ports. You can adopt WRR queue scheduling algorithm as required.

Configuration Procedure Follow these steps to configure WRR queues:

To do...	Use the command...	Remarks	
Enter system view	system-view	-	
Enter port view or port group view	Enter port view Enter port group view	interface <i>interface-type interface-number</i> port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	Perform either of the two operations. The configuration performed in Ethernet port view applies to the current port only. The configuration performed in port group view applies to all the ports in the port group
Configure WRR queue scheduling	qos wrr <i>queue-id</i> group <i>group-id</i> weight <i>schedule-value</i>	Required By default, all the ports adopt the WRR queue scheduling algorithm, with the weight values assigned to queue 0 through queue 7 being 1, 2, 3, 4, 5, 9, 13, and 15.	

Configuration Examples Network requirements

Configure WRR queue scheduling algorithm on GigabitEthernet1/0/1, and assign weight 1, 2, 4, 6, 8, 10, 12, and 14 to queue 0 through queue 7.

Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Configure the WRR queues on GigabitEthernet1/0/1 port.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] qos wrr 0 group 1 weight 1
[Sysname-GigabitEthernet1/0/1] qos wrr 1 group 1 weight 2
[Sysname-GigabitEthernet1/0/1] qos wrr 2 group 1 weight 4
[Sysname-GigabitEthernet1/0/1] qos wrr 3 group 1 weight 6
[Sysname-GigabitEthernet1/0/1] qos wrr 4 group 1 weight 8
[Sysname-GigabitEthernet1/0/1] qos wrr 5 group 1 weight 10
[Sysname-GigabitEthernet1/0/1] qos wrr 6 group 1 weight 12
[Sysname-GigabitEthernet1/0/1] qos wrr 7 group 1 weight 14
```

Configuring SP+WRR Queues

As required, you can configure part of the queues on the port to adopt the SP queue-scheduling algorithm and parts of queues to adopt the WRR queue-scheduling algorithm. Through adding the queues on a port to the SP scheduling group and WRR scheduling group (namely, group 1), the SP+WRR queue scheduling is implemented. During the queue scheduling process, the queues in the SP scheduling group is scheduled preferentially. When no packet is to be sent in the queues in the SP scheduling group, the queues in the WRR scheduling group are scheduled. The queues in the SP scheduling group are scheduled according to the strict priority of each queue, while the queues in the WRR queue scheduling group are scheduled according the weight value of each queue.

Configuration Procedure Follow these steps to configure SP + WRR queues:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter port view or port group view	interface <i>interface-type interface-number</i>	Perform either of the two operations.
Enter port group view	port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	The configuration performed in Ethernet port view applies to the current port only. The configuration performed in port group view applies to all the ports in the port group.
Configure SP queue scheduling	qos wrr <i>queue-id</i> group sp	Required By default, all the ports adopt the WRR queue scheduling algorithm, with the weight values assigned to queue 0 through queue 7 being 1, 2, 3, 4, 5, 9, 13, and 15.

To do...	Use the command...	Remarks
Configure WRR queue scheduling	qos wrr <i>queue-id</i> group <i>group-id</i> weight <i>queue-weight</i>	Required

Configuration Examples **Network requirements**

- Configure to adopt SP+WRR queue scheduling algorithm on GigabitEthernet1/0/1.
- Configure queue 0, queue 1, queue 2 and queue 3 on GigabitEthernet1/0/1 to be in SP queue scheduling group.
- Configure queue 4, queue 5, queue 6 and queue 7 on GigabitEthernet1/0/1 to be in WRR queue scheduling group, with the weight being 2, 4, 6 and 8 respectively.

Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Enable the SP+WRR queue scheduling algorithm on GigabitEthernet1/0/1.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] qos wrr 0 group sp
[Sysname-GigabitEthernet1/0/1] qos wrr 1 group sp
[Sysname-GigabitEthernet1/0/1] qos wrr 2 group sp
[Sysname-GigabitEthernet1/0/1] qos wrr 3 group sp
[Sysname-GigabitEthernet1/0/1] qos wrr 4 group 1 weight 2
[Sysname-GigabitEthernet1/0/1] qos wrr 5 group 1 weight 4
[Sysname-GigabitEthernet1/0/1] qos wrr 6 group 1 weight 6
[Sysname-GigabitEthernet1/0/1] qos wrr 7 group 1 weight 8
```

Displaying and Maintaining Congestion Management

To do...	Use the command...	Remarks
Display WRR queue configuration information	display qos wrr interface [<i>interface-type</i> <i>interface-number</i>]	Available in any view
Display SP queue configuration information	display qos sp interface [<i>interface-type</i> <i>interface-number</i>]	

69

PRIORITY MAPPING

When configuring priority mapping, go to these sections for information you are interested in:

- “Priority Mapping Overview” on page 883
- “Configuring a Priority Mapping Table” on page 884
- “Configuring the Port Priority” on page 885
- “Configuring Port Priority Trust Mode” on page 886
- “Displaying and Maintaining Priority Mapping” on page 887

Priority Mapping Overview

When a packet reaches a switch, the switch assigns the packet parameters according to its configuration, such as 802.1p precedence, DSCP precedence, IP precedence, local precedence, and drop precedence.

The local precedence and drop precedence are described as follows.

- Local precedence is the precedence that the switch assigns to a packet and it is corresponding to the number of an outbound queue on the port. Local precedence takes effect only on the local switch.
- Drop precedence is a parameter that is referred to when dropping packets. The higher the drop precedence, the more likely a packet is dropped.

The Switch 4800G provides the following two priority trust modes:

- Trusting the DSCP precedence of received packets. In this mode, the switch searches the **dscp-dot1p/dp/dscp** mapping table based on the DSCP precedence of the received packet for the 802.1p precedence/drop precedence/DSCP precedence to be used to mark the packet. Then the switch searches the **dot1p-lp** mapping table based on the marked 802.1p precedence for the corresponding local precedence and marks the received packet with the local precedence.
- Trusting the 802.1p precedence of received packets. In this mode, if a packet is received without an 802.1q tag, the switch takes the priority of the receiving port as the 802.1p precedence of the packet and then based on the priority searches the **dot1p-dp/lp** mapping table for the local/drop precedence for the packet. If a packet is received with an 802.1q tag, the switch searches the **dot1p-dp/lp** mapping table based on the 802.1p precedence in the tag for local/drop precedence for the packet.

The default **dot1p-lp/dp** mapping and **dscp-dot1p/dp/dscp** mapping provided by the Switch 4800G are shown in the following two tables.

Table 71 The default values of dot1p-lp mapping and dot1p-dp mapping

Imported priority value	dot1p-lp mapping	dot1p-dp mapping
802.1p precedence (dot1p)	Local precedence (lp)	Drop precedence (dp)
0	2	0
1	0	0
2	1	0
3	3	0
4	4	0
5	5	0
6	6	0
7	7	0

Table 72 The default values of dscp-dp mapping, dscp-dot1p mapping, and dscp-dscp mapping

Imported priority value	dscp-dp mapping	dscp-dot1p mapping	dscp-dscp mapping
DSCP precedence (dscp)	Drop precedence (dp)	802.1p precedence (dot1p)	DSCP precedence (dscp)
0 to 7	0	0	0
8 to 15	0	1	8
16 to 23	0	2	16
24 to 31	0	3	24
32 to 39	0	4	32
40 to 47	0	5	40
48 to 55	0	6	48
56 to 63	0	7	56



You cannot configure to map any DSCP value to drop precedence 1.

Configuring a Priority Mapping Table

You can modify the priority mapping tables in a switch as required.

Follow the two steps to configure priority mapping tables:

- Enter priority mapping table view;
- Configure priority mapping parameters.

Configuration Prerequisites

The new priority mapping table is determined.

Configuration Procedure

Follow these steps to configure a priority mapping table:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Enter priority mapping table view	qos map-table { dot1p-dp dot1p-lp dscp-dot1p dscp-dp dscp-dscp }	Required To configure a priority mapping table, you need to enter the corresponding priority mapping table view.
Configure priority mapping parameters	import <i>import-value-list</i> export <i>export-value</i>	Required The newly configured mapping entries overwrite the corresponding previous entries.

Configuration Examples Network requirements

Modify the **dot1p-lp** mapping table as those listed in Table 73.

Table 73 The specified dot1p-lp mapping

802.1p precedence	Local precedence
0	0
1	0
2	1
3	1
4	2
5	2
6	3
7	3

Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Enter **dot1p-lp** priority mapping table view.

```
[Sysname] qos map-table dot1p-lp
```

Modify **dot1p-lp** priority mapping parameters.

```
[Sysname-maptbl-dot1p-lp] import 0 1 export 0
[Sysname-maptbl-dot1p-lp] import 2 3 export 1
[Sysname-maptbl-dot1p-lp] import 4 5 export 2
[Sysname-maptbl-dot1p-lp] import 6 7 export 3
```

Configuring the Port Priority

By default, if a port receives packets without 802.1q tags, the switch takes the priority of the receiving port as the 802.1p precedence of the received packets, searches the **dot1p-lp/dp** mapping table for the corresponding local precedence and drop precedence according to the 802.1p precedence of the received packets, and then marks the received packets with the corresponding local precedence and drop precedence.

Port priority is in the range 0 to 7. You can set the port priority as required.

Configuration Prerequisites The port priority of the port is determined.

Configuration Procedure Follow these steps to configure port priority:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter port view or port group view	Enter port view interface <i>interface-type</i> <i>interface-number</i>	Perform either of the two operations.
	Enter port group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	The configuration performed in Ethernet port view applies to the current port only. The configuration performed in port group view applies to all the ports in the port group.
Configure port priority	qos priority <i>priority-value</i>	Required By default, the port priority is 0.

Configuration Examples **Network requirements**

Configure the port priority to 7.

Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Configure port priority of GigabitEthernet1/0/1.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] qos priority 7
```

Configuring Port Priority Trust Mode

You can configure the switch to trust the DSCP precedence of the received packets. In this case, the switch searches the **dscp-dot1p/dp/dscp** mapping table for the corresponding precedence according to the DSCP precedence of the packets and marks the received packets with the precedence.

Configuration Prerequisites It is determined to trust the DSCP precedence of received packets.

Configuration Procedure Follow these steps to configure the port priority trust mode:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter port view or port group view	Enter port view interface <i>interface-type</i> <i>interface-number</i>	Perform either of the two operations.
	Enter port group view port-group { manual <i>port-group-name</i> aggregation <i>agg-id</i> }	The configuration performed in Ethernet port view applies to the current port only. The configuration performed in port group view applies to all the ports in the port group.

To do...	Use the command...	Remarks
Configure to trust the DSCP precedence of the received packets	qos trust dscp	Required By default, the 802.1p precedence of the received packets is trusted.

Configuration Examples Network requirements

Configure to trust the DSCP precedence of the received packets.

Configuration procedure

Enter system view.

```
<Sysname> system-view
```

Enter port view.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1]
```

Configure to trust the DSCP precedence of the received packets.

```
[Sysname-GigabitEthernet1/0/1] qos trust dscp
```

Displaying and Maintaining Priority Mapping

To do...	Use the command...	Remarks
Display the information about a specified priority mapping table	display qos map-table [dot1p-dp dot1p-lp dscp-dot1p dscp-dp dscp-dscp]	Available in any view
Display the priority trust mode configured for a port	display qos trust interface [<i>interface-type interface-number</i>]	

70

APPLYING A QoS POLICY TO VLANs

When applying a QoS policy to VLANs, go to these sections for information that you are interested in:

- “Overview” on page 889
- “Applying a QoS Policy to VLANs” on page 889
- “Displaying and Maintaining QoS Policies Applied to VLANs” on page 890
- “Configuration Examples” on page 890

Overview

QoS polices support the following application modes:

- Port-based application: QoS policies are effective for inbound packets on a port.
- VLAN-based application: QoS policies are effective for inbound traffic on a VLAN.

A QoS policy is not effective on dynamic VLANs, for example, VLANs created by GVRP.

Applying a QoS Policy to VLANs

Configuration Prerequisites

- The QoS policy to be applied is defined. Refer to “Configuring a QoS Policy” on page 870 for policy defining.
- VLANs where the QoS policy is to be applied are determined.

Configuration Procedure

Follow these steps to apply a QoS policy to VLANs:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Apply the QoS policy to the specified VLAN(s)	qos vlan-policy <i>policy-name</i> vlan <i>vlan-id-list</i> { inbound outbound }	Required

Note that, when you apply a QoS policy with the **qos vlan-policy** command, the support for the **inbound/outbound** keyword varies with the actions defined in the traffic behavior, as described in Table 70.

Displaying and Maintaining QoS Policies Applied to VLANs

To do...	Use the command...	Remarks
Display the QoS policies applied to VLANs	display qos vlan-policy { name <i>policy-name</i> vlan [<i>vlan-id</i>] }	Available in any view
Clear the statistics information about the QoS policies applied to VLANs	reset qos vlan-policy [vlan <i>vlan-id</i>]	Available in user view

Configuration Examples

Network Requirements

- The QoS policy **test** is defined to perform traffic policing for the packets matching basic IPv4 ACL 2000, with CIR as 64 kbps. The exceeding packets are dropped.
- Apply the VLAN policy **test** to the inbound direction of VLAN 200, VLAN 300, VLAN 400, VLAN 500, VLAN 600, VLAN 700, VLAN 800, and VLAN 900.

Configuration Procedure

Enter system view.

```
<Sysname> system-view
```

Create a class and enter class view.

```
[Sysname] traffic classifier cl1
```

Define a classification rule.

```
[Sysname-classifier-cl1] if-match acl 2000
[Sysname-classifier-cl1] quit
```

Create a traffic behavior and enter traffic behavior view.

```
[Sysname] traffic behavior be1
```

Configure the traffic behavior.

```
[Sysname-behavior-be1] car cir 64
[Sysname-behavior-be1] quit
```

Create a QoS policy and enter QoS policy view.

```
[Sysname] qos policy test
```

Associate a class with a traffic behavior.

```
[Sysname-qospolicy-test] classifier cl1 behavior be1
[Sysname-qospolicy-test] quit
```

Apply the policy to specific VLANs.

```
[Sysname] qos vlan-policy test vlan 200 300 400 500 600 700 800 900 inbound
```

When configuring traffic mirroring, go to these sections for information that you are interested in:

- “Overview” on page 891
- “Configuring Traffic Mirroring” on page 891
- “Displaying and Maintaining Traffic Mirroring” on page 892
- “Traffic Mirroring Configuration Examples” on page 892

Overview

Traffic mirroring is to replicate the specified packets to the specified destination. It is generally used for testing and troubleshooting the network.

Depending on different types of mirroring destinations, there are three types of traffic mirroring:

- Mirroring to port: The desired traffic on a mirrored port is replicated and sent to a destination port (that is, a mirroring port).
- Mirroring to CPU: The desired traffic on a mirrored port is replicated and sent to the CPU on the module of the port for further analysis.
- Mirroring to VLAN: The desired traffic on a mirrored port is replicated and sent to a VLAN, where the traffic is broadcast and all the ports (if available) in the VLAN will receive the traffic. If the destination VLAN does not exist, you can still configure the function, and the function will automatically take effect after the VLAN is created and a port is added to it.



On the Switch 4800G, traffic can only be mirrored to ports and to CPU.

Configuring Traffic Mirroring

To configure traffic mirroring, you must enter the view of an existing traffic behavior.

Follow these steps to configure traffic mirroring to a port:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter traffic behavior view	traffic behavior <i>behavior-name</i>	Required
Configure traffic mirroring action in the traffic behavior	mirror-to { cpu interface <i>interface-type</i> <i>interface-number</i> }	Required

Displaying and Maintaining Traffic Mirroring

To do...	Use the command...	Remarks
Display the configuration information about the user-defined traffic behavior	display traffic behavior user-defined <i>behavior-name</i>	Available in any view
Display the configuration information about the user-defined policy	display qos policy user-defined <i>policy-name</i>	

Traffic Mirroring Configuration Examples

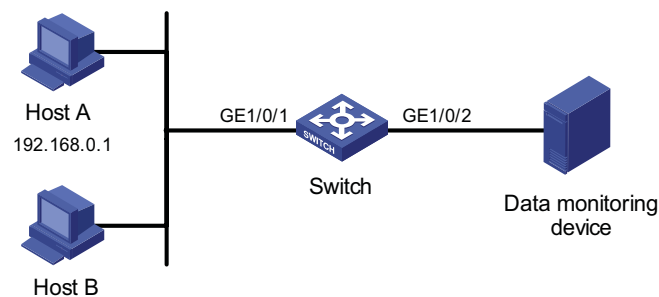
Network Requirements

The user's network is as described below:

- Host A (with the IP address 192.168.0.1) and Host B are connected to GigabitEthernet1/0/1 of the switch.
- The data monitoring device is connected to GigabitEthernet1/0/2 of the switch.

It is required to monitor and analyze packets sent by Host A on the data monitoring device.

Figure 262 Network diagram for configuring traffic mirroring to a port



Configuration Procedure

Configure Switch:

Enter system view.

```
<Sysname> system-view
```

Configure basic IPv4 ACL 2000 to match packets with the source IP address 192.168.0.1.

```
[Sysname] acl number 2000
[Sysname-acl-basic-2000] rule permit source 192.168.0.1 0
[Sysname-acl-basic-2000] quit
```

Configure a traffic classification rule to use ACL 2000 for traffic classification.

```
[Sysname] traffic classifier 1
[Sysname-classifier-1] if-match acl 2000
[Sysname-classifier-1] quit
```


Configure a traffic behavior and define the action of mirroring traffic to GigabitEthernet1/0/2 in the traffic behavior.

```
[Sysname] traffic behavior 1
[Sysname-behavior-1] mirror-to interface GigabitEthernet 1/0/2
[Sysname-behavior-1] quit
```

Configure a QoS policy and associate traffic behavior 1 with classification rule 1.

```
[Sysname] qos policy 1
[Sysname-policy-1] classifier 1 behavior 1
[Sysname-policy-1] quit
```

Apply the policy in the inbound direction of GigabitEthernet1/0/1.

```
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet1/0/1] qos apply policy 1 inbound
```

After the configurations, you can monitor all packets sent from Host A on the data monitoring device.

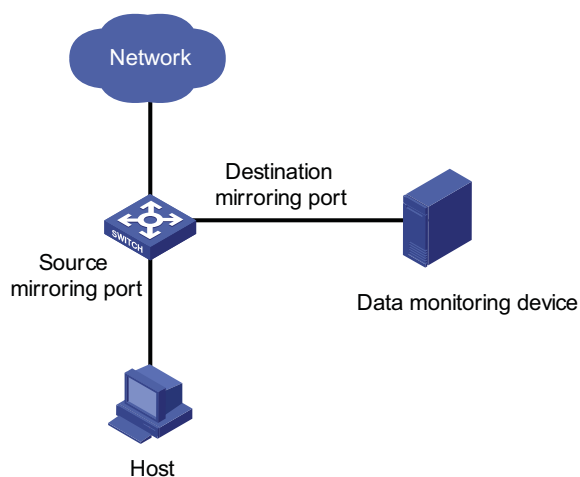
When configuring port mirroring, go to these sections for information you are interested in:

- "Introduction to Port Mirroring" on page 895
- "Configuring Local Port Mirroring" on page 897
- "Configuring Remote Port Mirroring" on page 898
- "Displaying and Maintaining Port Mirroring" on page 899
- "Port Mirroring Configuration Examples" on page 900

Introduction to Port Mirroring

Port mirroring allows you to duplicate the packets passing specified ports to the destination mirroring port. As destination mirroring ports usually have data monitoring devices connected to them, you can analyze the packets duplicated to the destination mirroring port on these devices so as to monitor and troubleshoot the network.

Figure 263 A port mirroring implementation



Classification of Port Mirroring

There are two kinds of port mirroring: local port mirroring and remote port mirroring.

- Local port mirroring copies packets passing through one or more ports (known as source ports) of a device to the monitor port (also destination port) for analysis and monitoring purpose. In this case, the source ports and the destination port are located on the same device.
- Remote port mirroring implements port mirroring between multiple devices. That is, the source ports and the destination port can be located on different

devices in a network. Currently, remote port mirroring can only be implemented on Layer 2.

Implementing Port Mirroring

Port mirroring is implemented through port mirroring groups, which fall into these three categories: local port mirroring group, remote source port mirroring group, and remote destination port mirroring group. Two port mirroring implementation modes are introduced in the following section.

Local port mirroring

Local port mirroring is implemented by local port mirroring group.

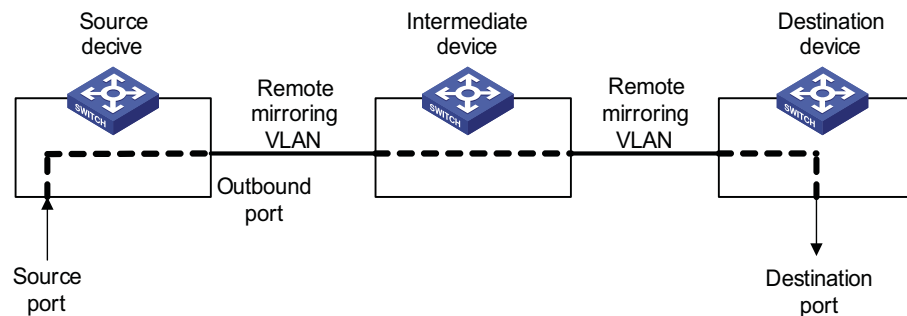
In this mode, the source ports and the destination port are in the same local port mirroring group. Packets passing through the source ports are duplicated and then forwarded to the destination port.

Remote port mirroring

Remote port mirroring is achieved through the cooperation of remote source port mirroring group and remote destination port mirroring group.

Figure 264 illustrates a remote port mirroring implementation.

Figure 264 A remote mirroring implementation



The devices in Figure 264 function as follows:

- Source device

Source device contains source mirroring ports, and remote source port mirroring groups are created on source devices. A source device duplicates the packets passing the source ports on it and sends them to the outbound port. The packets are then broadcast in the remote mirroring VLAN and are received by the intermediate device or destination device.

- Intermediate device

Intermediate devices are used to connect source devices and destination devices. An intermediate device forwards the mirrored packets to the next intermediate device or the destination device. If the source device is directly connected to the destination device, no intermediate device is needed. In a remote mirroring VLAN, the source devices and the destination device need to be able to communicate with one another on Layer 2.

- Destination device

Destination device contains destination mirroring port, and remote destination port mirroring groups are created on destination devices. Upon receiving a mirrored packet, the destination device checks to see if the VLAN ID of the received packet is the same as that of the remote mirroring VLAN of the remote destination port mirroring group. If yes, the destination device forwards the packet to the monitoring device through the destination mirroring port.

Other Functions Supported by Port Mirroring

In addition, in a port mirroring group, a destination port can monitor multiple source ports simultaneously in the mirroring group.

Configuring Local Port Mirroring

Follow these steps to configure local port mirroring:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Create a local mirroring group		mirroring-group <i>group-id</i> local	Required
Add ports to the port mirroring group as source ports	In system view	mirroring-group <i>group-id</i> mirroring-port <i>mirroring-port-list</i> { both inbound outbound }	Use either approach. You can add ports to a port mirroring group as source ports in either system view or interface view.
	In interface view	interface <i>interface-type</i> <i>interface-number</i> [mirroring-group <i>group-id</i>] mirroring-port { both inbound outbound } quit	In system view, you can add multiple ports to a port mirroring group at one time. While in interface view, you can only add the current port to a port mirroring group.
Add a port to the mirroring group as the destination port	In system view	mirroring-group <i>group-id</i> monitor-port <i>monitor-port-id</i>	Use either approach. You can add a destination port to a port mirroring group in either system view or interface view. They achieve the same purpose.
	In interface view	interface <i>interface-type</i> <i>interface-number</i> [mirroring-group <i>group-id</i>] monitor-port	



- *A local mirroring group is effective only when it has both source ports and the destination port configured.*
- *It is not recommended to enable STP, RSTP or MSTP on the destination port; otherwise, the mirroring function may be affected.*
- *Do not use the destination mirroring port for any purpose other than port mirroring.*
- *The source ports and the destination port cannot be the member ports of the current mirroring group.*
- *Before adding the destination port for a port mirroring group, make sure the port mirroring group exists. A mirroring group can have only one destination port.*

Configuring Remote Port Mirroring

Configuring a Remote Source Mirroring Group

Follow these steps to configure a remote port mirroring group

To do...		Use the command...	Remarks
Enter system view		system-view	-
Create a remote source mirroring group		mirroring-group <i>group-id</i> remote-source	Required
Add ports to the mirroring group as source ports	In system view	mirroring-group <i>group-id</i> mirroring-port <i>mirroring-port-list</i> { both inbound outbound }	Use either approach. You can add ports to a source port mirroring group in either system view or interface view. They achieve the same purpose.
	In interface view	interface <i>interface-type</i> <i>interface-number</i> [mirroring-group <i>group-id</i>] mirroring-port { both inbound outbound } quit	
Add a port to the mirroring group as the outbound mirroring port	In system view	mirroring-group <i>group-id</i> monitor-egress <i>monitor-egress-port-id</i>	Use either approach. You can add ports to a source mirroring group in either system view or interface view. They achieve the same purpose.
	In interface view	interface <i>interface-type</i> <i>interface-number</i> mirroring-group <i>group-id</i> monitor-egress quit	
Configure the remote port mirroring VLAN for the mirroring group		mirroring-group <i>group-id</i> remote-probe vlan <i>rprobe-vlan-id</i>	Required



- *All ports in a remote mirroring group belong to the same device. A remote source mirroring group can have only one outbound mirroring port.*
- *The outbound mirroring port cannot be a member port of the current mirroring group.*
- *It is not recommended to add the source ports to a remote VLAN, which can be used for remote mirroring only.*
- *It is not recommended to configure STP, RSTP, MSTP, 802.1x, IGMP Snooping, static ARP and MAC address learning on the outbound mirroring port; otherwise, the mirroring function may be affected.*
- *Only existing static VLANs can be configured as remote port mirroring VLANs. To remove a VLAN operating as a remote port mirroring VLAN, you need to restore it to a normal VLAN first. A remote port mirroring group gets invalid if the corresponding remote port mirroring VLAN is removed.*
- *A port can belong to only one port mirroring group. A VLAN can be the remote port mirroring VLAN of only one port mirroring group.*

Configuring a Remote Destination Port Mirroring Group

Follow these steps to configure a remote destination port mirroring group:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Create a remote destination port mirroring group		mirroring-group <i>group-id</i> remote-destination	Required
Configure the remote port mirroring VLAN for the port mirroring group		mirroring-group <i>group-id</i> remote-probe vlan <i>rprobe-vlan-id</i>	Required
Add a port to the port mirroring group as the destination port	In system view In interface view	mirroring-group <i>group-id</i> monitor-port <i>monitor-port-id</i> interface <i>interface-type</i> <i>interface-number</i> [mirroring-group <i>group-id</i>] monitor-port quit	Use either approach. You can add a port to a remote port mirroring group as the destination port in either system view or interface view. They achieve the same purpose.
Enter destination interface view		interface <i>interface-type</i> <i>interface-number</i>	-
Add the port to the remote port mirroring VLAN	The port is an access port The port is a trunk port The port is a hybrid port	port access vlan <i>rprobe-vlan-id</i> port trunk permit vlan <i>rprobe-vlan-id</i> port hybrid vlan <i>rprobe-vlan-id</i> { tagged untagged }	Perform one of these three operations according to the port type.



- *The remote destination mirroring port cannot be a member port of the current mirroring group.*
- *The remote destination mirroring port can be an access, trunk, or hybrid port. It must be assigned to the remote mirroring VLAN.*
- *Do not enable STP, RSTP or MSTP on the remote destination mirroring port. Otherwise, the mirroring function may be affected.*
- *Do not use the remote destination mirroring port for any purpose other than port mirroring.*
- *Only existing static VLANs can be configured as remote port mirroring VLANs. To remove a VLAN operating as a remote port mirroring VLAN, you need to restore it to a normal VLAN first. A remote port mirroring group gets invalid if the corresponding remote port mirroring VLAN is removed.*
- *Use a remote port mirroring VLAN for remote port mirroring only.*
- *A port can belong to only one port mirroring group. A VLAN can be the remote port mirroring VLAN of only one port mirroring group.*

Displaying and Maintaining Port Mirroring

To do...	Use the command...	Remarks
Display the configuration of a port mirroring group	display mirroring-group { <i>group-id</i> all local remote-destination remote-source }	Available in any view

Port Mirroring Configuration Examples

Local Port Mirroring Configuration Example

Network requirements

The departments of a company connect to each other through Ethernet switches:

- Research and Development (R&D) department is connected to Switch C through GigabitEthernet 1/0/1.
- Marketing department is connected to Switch C through GigabitEthernet 1/0/2.
- Data monitoring device is connected to Switch C through GigabitEthernet 1/0/3

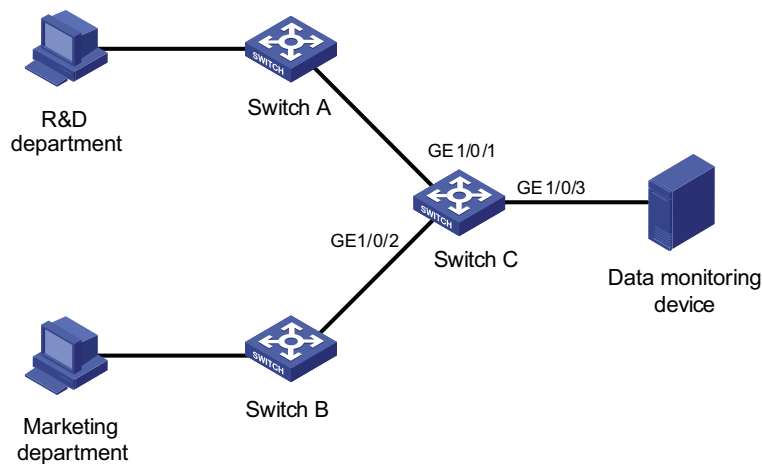
The administrator wants to monitor the packets received on and sent from the R&D department and the marketing department through the data monitoring device.

Use the local port mirroring function to meet the requirement. Perform the following configurations on Switch C.

- Configure GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 as mirroring source ports.
- Configure GigabitEthernet 1/0/3 as the mirroring destination port.

Network diagram

Figure 265 Network diagram for local port mirroring configuration



Configuration procedure

Configure Switch C.

Create a local port mirroring group.

```

<SwitchC> system-view
[SwitchC] mirroring-group 1 local
  
```


Add port GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to the port mirroring group as source ports. Add port GigabitEthernet 1/0/3 to the port mirroring group as the destination port.

```
[SwitchC] mirroring-group 1 mirroring-port GigabitEthernet 1/0/1 GigabitEthernet 1/0/2 both
[SwitchC] mirroring-group 1 monitor-port GigabitEthernet 1/0/3
```

Display the configuration of all the port mirroring groups.

```
[SwitchC] display mirroring-group all
mirroring-group 1:
  type: local
  status: active
  mirroring port:
    GigabitEthernet1/0/1 both
    GigabitEthernet1/0/2 both
  monitor port: GigabitEthernet1/0/3
```

After finishing the configuration, you can monitor all the packets received and sent by R&D department and Marketing department on the Data monitoring device.

Remote Port Mirroring Configuration Example

Network requirements

The departments of a company connect to each other through Ethernet switches:

- Department 1 is connected to GigabitEthernet 1/0/1 of Switch A.
- Department 2 is connected to GigabitEthernet 1/0/2 of Switch A.
- GigabitEthernet 1/0/3 of Switch A connects to GigabitEthernet 1/0/1 of Switch B.
- GigabitEthernet 1/0/2 of Switch B connects to GigabitEthernet 1/0/1 of Switch C.
- The data monitoring device is connected to GigabitEthernet 1/0/2 of Switch C.

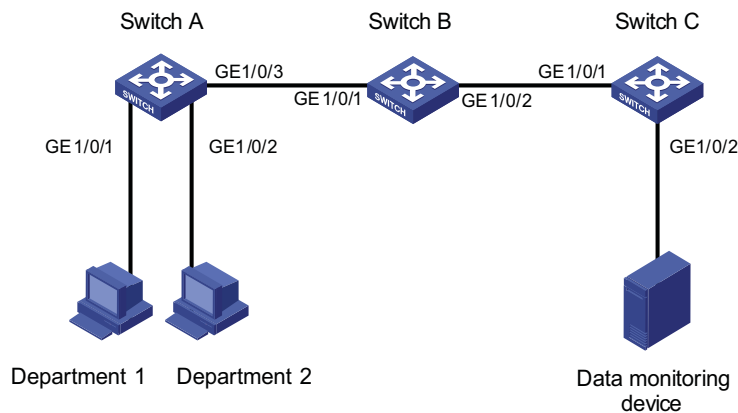
The administrator wants to monitor the packets sent from Department 1 and 2 through the data monitoring device.

Use the remote port mirroring function to meet the requirement. Perform the following configurations:

- Use Switch A as the source device, Switch B as the intermediate device, and Switch C as the destination device.
- On Switch A, create a remote source mirroring group; create VLAN 2 and configure it as the remote port mirroring VLAN; add port GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to the port mirroring group as two source ports. Configure port GigabitEthernet 1/0/3 as the outbound mirroring port.
- Configure port GigabitEthernet 1/0/3 of Switch A, port GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 of Switch B, and port GigabitEthernet 1/0/1 of Switch C as trunk ports and configure them to permit packets of VLAN 2.
- Create a remote destination mirroring group on Switch C. Configure VLAN 2 as the remote port mirroring VLAN and port GigabitEthernet 1/0/2, to which the data monitoring device is connected, as the destination port.

Network diagram

Figure 266 Network diagram for remote port mirroring configuration



Configuration procedure

1 Configure Switch A (the source device).

Create a remote source port mirroring group.

```
<SwitchA> system-view
[SwitchA] mirroring-group 1 remote-source
```

Create VLAN 2.

```
[SwitchA] vlan 2
[SwitchA-vlan2] quit
```

Configure VLAN 2 as the remote port mirroring VLAN of the remote port mirroring group. Add port GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 to the remote port mirroring group as source ports. Configure port GigabitEthernet 1/0/3 as the outbound mirroring port.

```
[SwitchA] mirroring-group 1 remote-probe vlan 2
[SwitchA] mirroring-group 1 mirroring-port GigabitEthernet 1/0/1 GigabitEthernet 1/0/2 inbound
[SwitchA] mirroring-group 1 monitor-egress GigabitEthernet 1/0/3
```

Configure port GigabitEthernet 1/0/3 as a trunk port and configure the port to permit the packets of VLAN 2.

```
[SwitchA] interface GigabitEthernet 1/0/3
[SwitchA-GigabitEthernet1/0/3] port link-type trunk
[SwitchA-GigabitEthernet1/0/3] port trunk permit vlan 2
```

1 Configure Switch B (the intermediate device).

Configure port GigabitEthernet 1/0/1 as a trunk port and configure the port to permit the packets of VLAN 2.

```
<SwitchB> system-view
[SwitchB] interface GigabitEthernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] port link-type trunk
[SwitchB-GigabitEthernet1/0/1] port trunk permit vlan 2
[SwitchB-GigabitEthernet1/0/1] quit
```

Configure port GigabitEthernet 1/0/2 as a trunk port and configure the port to permit the packets of VLAN 2.

```
[SwitchB] interface GigabitEthernet 1/0/2
[SwitchB-GigabitEthernet1/0/2] port link-type trunk
[SwitchB-GigabitEthernet1/0/2] port trunk permit vlan 2
```

1 Configure Switch C (the destination device).

Configure port GigabitEthernet 1/0/1 as a trunk port and configure the port to permit the packets of VLAN 2.

```
<SwitchC> system-view
[SwitchC] interface GigabitEthernet 1/0/1
[SwitchC-GigabitEthernet1/0/1] port link-type trunk
[SwitchC-GigabitEthernet1/0/1] port trunk permit vlan 2
[SwitchC-GigabitEthernet1/0/1] quit
```

Create a remote destination port mirroring group.

```
[SwitchC] mirroring-group 1 remote-destination
```

Create VLAN 2.

```
[SwitchC] vlan 2
[SwitchC-vlan2] quit
```

Configure VLAN 2 as the remote port mirroring VLAN of the remote destination port mirroring group. Add port GigabitEthernet 1/0/2 to the remote destination port mirroring group as the destination port.

```
[SwitchC] mirroring-group 1 remote-probe vlan 2
[SwitchC] mirroring-group 1 monitor-port GigabitEthernet 1/0/2
[SwitchC] interface GigabitEthernet 1/0/2
[SwitchC-GigabitEthernet1/0/2] port access vlan 2
```

After finishing the configuration, you can monitor all the packets sent by Department 1 and Department 2 on the Data monitoring device.

CLUSTER MANAGEMENT CONFIGURATION

When configuring cluster management, go to these sections for information you are interested in:

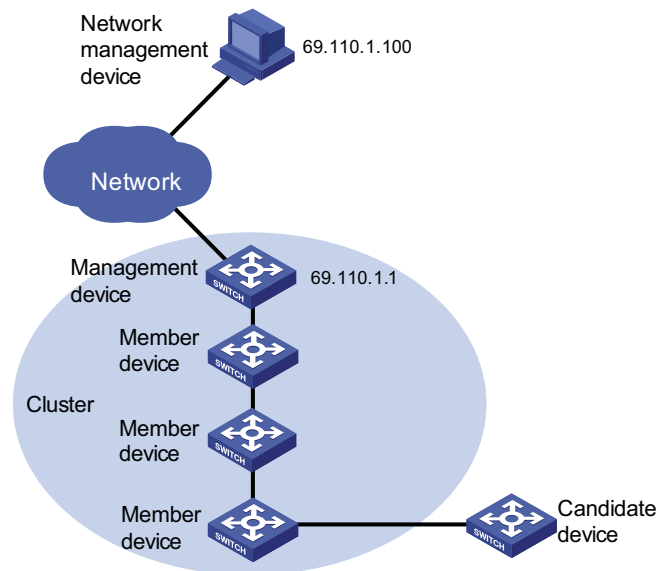
- "Cluster Management Overview" on page 905
- "Cluster Configuration Task List" on page 911
- "Configuring the Management Device" on page 912
- "Configuring the Member Devices" on page 917
- "Configuring Access Between the Management Device and Its Member Devices" on page 918
- "Adding a Candidate Device to a Cluster" on page 919
- "Configuring Advanced Cluster Functions" on page 919
- "Displaying and Maintaining Cluster Management" on page 922
- "Cluster Management Configuration Examples" on page 922

Cluster Management Overview

Cluster Management Definition

A cluster is an aggregation of a group of communication devices. Cluster management is to implement management of large numbers of distributed network devices.

Cluster management is implemented through 3Com Group Management Protocol version 2 (Switch Clusteringv2). By employing Switch Clusteringv2, a network administrator can manage multiple devices using the public IP address of one device in a cluster. The device that configured with a public address and performs the management function is known as the management device and other managed devices are called member devices, which together form a cluster. Figure 267 illustrates a typical cluster implementation.

Figure 267 Network diagram for a cluster

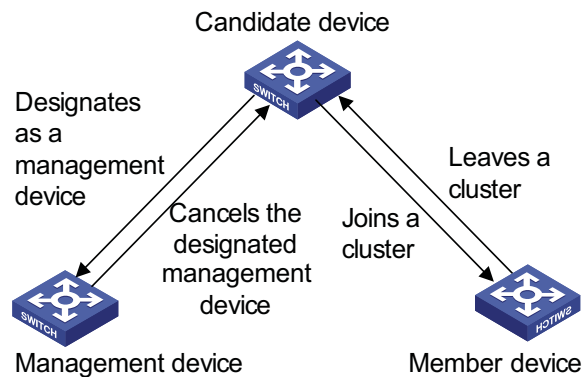
Cluster management offers the following advantages:

- Saving public IP address resource
- Simplifying configuration and management tasks. By configuring a public IP address on the management device, you can configure and manage a group of member devices on the management device without the trouble of logging onto each device separately.
- Providing topology discovery and display function, which is useful for network monitoring and debugging
- Allowing simultaneous software upgrading and parameter configuring on multiple devices, free of topology and distance limitations

Roles in a Cluster

The devices in a cluster play different roles according to their different functions and status. You can specify the role a device plays. The following three roles exist in a cluster: management device, member device, and candidate device.

- **Management device:** The device providing management interfaces for all devices in the cluster and the only device configured with a public IP address. Any configuration, management, and monitoring of the member devices in a cluster can only be implemented through the management device. When a device is specified as the management device, it collects Neighbor Discovery Protocol (NDP) and Neighbor Topology Discovery Protocol (NTDP) information to discover and define a candidate device.
- **Member device:** The device being managed by the management device in a cluster.
- **Candidate device:** A device that does not belong to any cluster but can be added to a cluster. Different from a member device, its topology information has been collected by the management device but it has not been added to the cluster.

Figure 268 Role change in a cluster

A device in a cluster changes its role according to the following rules:

- A candidate device becomes a management device when you create a cluster on it. Note that a cluster must have one (and only one) management device. On becoming a management device, the device collects network topology information and tries to discover and determine candidate devices, which can then be added to the cluster through configuration.
- A candidate device becomes a member device after being added to a cluster.
- A member device becomes a candidate device after it is removed from the cluster.
- A management device becomes a candidate device only after the cluster is removed.

How a Cluster Works

Switch Clusteringv2 consists of the following three protocols:

- Neighbor Discovery Protocol (NDP)
- Neighbor Topology Discovery Protocol (NTDP)
- Cluster

A cluster configures and manages the devices in it through the above three protocols.

Cluster management involves topology information collection and the establishment and maintenance of a cluster. Topology information collection and cluster maintenance are independent from each other, with the former starting before the cluster is created:

- All devices use NDP to collect the information of the directly connected neighbors, including their software version, host name, MAC address and port number.
- The management device uses NTDP to collect the information of the devices within user-specified hops and the topology information of all devices and specify the candidate devices of the cluster.
- The management device adds or deletes a member device and modifies cluster management configuration according to the candidate device information collected through NTDP.

Introduction to NDP

NDP is used to discover the information about directly connected neighbors, including the device name, software version, and connecting port of the adjacent devices. NDP works in the following ways:

- A device running NDP periodically sends NDP packets to its neighbors. An NDP packet carries NDP information (including the device name, software version, and connecting port, etc.) and the holdtime, which indicates how long the receiving devices will keep the NDP information. At the same time, the device also receives but does not forward the NDP packets from its neighbors.
- A device running NDP stores and maintains an NDP table. The device creates an entry in the NDP table for each neighbor. If a new neighbor is found, meaning the device receives an NDP packet sent by the neighbor for the first time, the device adds an entry in the NDP table. When another NDP packet is received, if the NDP information carried in the NDP packet is different from the stored information, the corresponding entry in the NDP table is updated; otherwise, only the holdtime of the entry is updated. If no NDP information from the neighbor is received within the holdtime, the corresponding entry is removed from the NDP table.

NDP runs on the data link layer, and therefore supports different network layer protocols.

Introduction to NTDP

NTDP is a protocol used to collect network topology information. NTDP provides information required for cluster management: it collects topology information about the devices within the specified hop count, to identify candidate devices for a cluster.

Based on the neighbor information stored in the neighbor table maintained by NDP, NTDP on the management device advertises NTDP topology collection requests to collect the NDP information of each device in a specific network range as well as the connection information of all its neighbors. The information collected will be used by the management device or the network management software to implement required functions.

When a member device detects a change on its neighbors through its NDP table, it informs the management device through handshake packets. Then the management device triggers its NTDP to perform specific topology collection, so that its NTDP can discover topology changes timely.

The management device collects topology information periodically. You can also administratively launch a topology information collection with commands. The process of topology information collection is as follows:

- The management device periodically sends NTDP topology collection request from the NTDP-enabled ports.
- Upon receiving the request, the device sends NTDP topology collection response to the management device, copies this response packet on the NTDP-enabled port and sends it to the adjacent device. Topology collection response includes the basic information of the NDP-enabled device and NDP information of all adjacent devices.

- The adjacent device performs the same operation until the NTDP topology collection request is sent to all the devices within specified hops.

When the NTDP topology collection request is advertised in the network, large numbers of network devices receive the NTDP topology collection request and send NTDP topology collection response at the same time, which may cause congestion and the management device busyness. To avoid such case, the following methods can be used to control the speed of the NTDP topology collection request advertisement:

- Upon receiving an NTDP topology collection request the device does not forward it, instead, it waits for a period of time and then forwards the NTDP topology collection request on the first NTDP-enabled port.
- On the same device, except the first port, each NTDP-enabled port waits for a period of time and then forwards the NTDP topology collection request after the port before it sends the NTDP topology collection request.

Cluster management maintenance

1 Adding a candidate device to a cluster

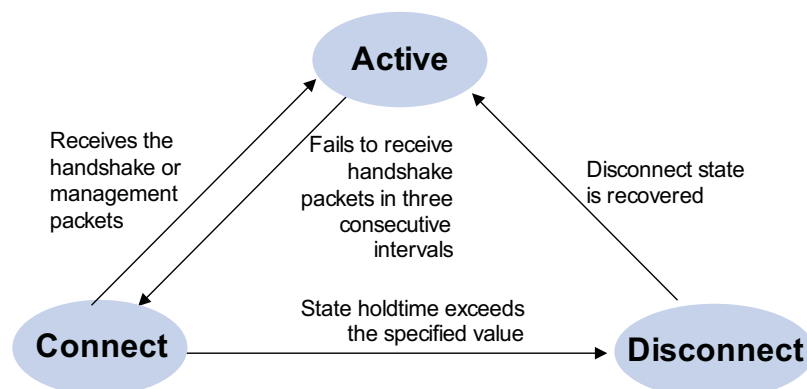
You should specify the management device before creating a cluster. The management device discovers and defines a candidate device through NDP and NTDP protocols. The candidate device can be automatically or manually added to the cluster.

After the candidate device is added to the cluster, it can obtain the member number assigned by the management device and the private IP address used for cluster management.

2 Communication within a cluster

In a cluster the management device communicates with its member devices by sending handshake packets to maintain connection between them. The management/member device state change is shown in Figure 269.

Figure 269 Management/member device state change



- After a cluster is created and a candidate device is added to the cluster and becomes a member device, the management device saves the state information of its member device and identifies it as Active. And the member device also saves its state information and identifies it as Active.
- After a cluster is created, its member devices begin to send handshake packets first. The management device also sends handshake packets to the member

devices at the same interval. Upon receiving the handshake packets from the other side, the management device or member device simply changes or remains its state as Active, without sending a response.

- If the management device does not receive the handshake packets from a member device in an interval three times of the interval to send handshake packets, it changes the status of the member device from Active to Connect. Likewise, if a member device fails to receive the handshake packets from the management device in an interval three times of the interval to send handshake packets, the status of the member device will also be changed from Active to Connect.
- If this management device, in information holdtime, receives the handshake or management packets from its member device which is in Connect state, it changes the state of its member device to Active; otherwise, it changes the state of its member device to Disconnect, in which case the management device considers its member device disconnected. If this member device, which is in Connect state, receives handshake or management packets from the management device in information holdtime, it changes its state to Active; otherwise, it changes its state to Disconnect.
- If the communication between the management device and a member device is recovered, the member device which is in Disconnect state will be added to the cluster. After that, the state of the member device locally and on the management device will be changed to Active.

Besides, the member device informs the management device using handshake packets when there is a neighbor topology change.

Management VLAN

The management VLAN limits the cluster management range. Through configuration of the management VLAN, the following functions can be implemented:

- Management packets (including NDP, NTDP and handshake packets) are restricted within the management VLAN, therefore isolated from other packets, which enhances security.
- The management device and the member devices communicate with each other through the management VLAN.

For a cluster to work normally, you must set the packets from the management VLAN to pass the subtending ports (If a candidate device is connected to the management device through another candidate device, the ports connecting these two candidate devices are called subtending ports.) and the ports connecting the management device and the member/candidate devices. Therefore:

- If the packets from the management VLAN cannot pass a port, the device connected with the port cannot be added to the cluster. Therefore, if the ports (including the subtending ports) connecting the management device and the member/candidate devices prohibit the packets from the management VLAN, you can set the packets from the management VLAN to pass the ports on candidate devices with the management VLAN auto-negotiation function.
- Only when the default VLAN ID of the subtending ports and the ports connecting the management device and the member/candidate devices is that

of the management VLAN can you set the packets without tags from the management VLAN to pass the ports; otherwise, only the packets with tags from the management VLAN can pass the ports.

Refer to “Introduction to VLAN” on page 83.

Cluster Configuration Task List

Before configuring a cluster, you need to determine the roles and functions the devices play. You also need to configure the related functions, preparing for the communication between devices within the cluster.

Complete these tasks to configure a cluster:

Tasks	Remarks	
“Configuring the Management Device” on page 912	“Enabling NDP Globally and for Specific Ports” on page 912	Optional
	“Configuring NDP Parameters” on page 912	Optional
	“Enabling NTDP Globally and for Specific Ports” on page 912	Optional
	“Configuring NTDP Parameters” on page 913	Optional
	“Manually Collecting NTDP Information” on page 913	Optional
	“Enabling the Cluster Function” on page 914	Optional
	“Establishing a Cluster” on page 914	Required
	“Configuring Communication Between the Management Device and the Member Devices Within a Cluster” on page 916	Optional
	“Configuring the Destination MAC Address of Cluster Management Multicast Packets” on page 916	Optional
	“Configuring Cluster Member Management” on page 916	Optional
“Configuring the Member Devices” on page 917	“Enabling NDP Globally and for Specific Ports” on page 917	Optional
	“Enabling NTDP Globally and for Specific Ports” on page 917	Optional
	“Manually Collecting NTDP Information” on page 917	Optional
	“Enabling the Cluster Function” on page 917	Optional
	“Deleting a Member Device from a Cluster” on page 917	Optional
“Configuring Access Between the Management Device and Its Member Devices” on page 918	Optional	
“Adding a Candidate Device to a Cluster” on page 919	Optional	
“Configuring Advanced Cluster Functions” on page 919	“Configuring Topology Management” on page 919	Optional
	“Configuring Interaction for a Cluster” on page 920	Optional



CAUTION: Disabling the NDP and NTDP functions on the management device and member devices after a cluster is created will not cause the cluster to be dismissed, but will influence the normal operation of the cluster.

Configuring the Management Device

Enabling NDP Globally and for Specific Ports

Follow these steps to enable NDP globally and for specific ports:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable NDP globally	ndp enable	Optional Enabled by default.
Enable the NDP feature for the port(s)	In system view ndp enable interface interface-list In Ethernet port view interface interface-type interface-number ndp enable	Use either command By default, NDP is enabled globally and also on all ports.



CAUTION:

- For NDP to work normally, you must enable NDP both globally and on the specified port.
- If the subtending port or the port connecting the management device to a member/candidate device is a port of a member in an aggregation group, you must enable NDP on all member ports of the aggregation group at the same time. Otherwise, NDP will work abnormally.
- You are recommended to disable NDP on the port which connects with the devices that do not need to join the cluster, preventing the management device from adding the device which needs not to join the cluster and collecting the topology information of this device.

Configuring NDP Parameters

Follow these steps to configure NDP parameters:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the period for the receiving devices to keep the NDP packets	ndp timer aging aging-time	Optional 180 seconds by default.
Configure the interval to send NDP packets	ndp timer hello hello-time	Optional 60 seconds by default.



CAUTION: The time for the receiving device to hold NDP packets cannot be shorter than the interval to send NDP packets; otherwise, the NDP table may become instable.

Enabling NTDP Globally and for Specific Ports

Follow these steps to enable NTDP globally and for specific ports:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable NTDP globally	ntdp enable	Optional Enabled by default
Enable NTDP for the port	interface <i>interface-type</i> <i>interface-number</i> ntdp enable	Optional NTDP is enabled on all ports by default.

**CAUTION:**

- For NTDP to work normally, you must enable NTDP both globally and on the specified port.
- The NTDP function is mutually exclusive with the BPDU TUNNEL function under a port and you cannot enable them at the same time. For the detailed description of the BPDU TUNNEL function, refer to “BPDU Tunneling Configuration” on page 141.
- If the subtending port or the port connecting the management device to a member/candidate device is a port of a member in an aggregation group, you must enable NDP on all member ports of the aggregation group at the same time. Otherwise, NDP will work abnormally.
- You are recommended to disable NDP on the port which connects with the devices that do not need to join the cluster, preventing the management device from adding the device which needs not to join the cluster and collecting the topology information of this device.

Configuring NTDP Parameters

Follow these steps to configure NTDP parameters:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the range within which topology information is to be collected	ntdp hop <i>hop-value</i>	Optional By default, the hop range for topology collection is 3 hops.
Configure the interval to collect topology information	ntdp timer <i>interval-time</i>	Optional 1 minute by default.
Configure the delay to forward topology-collection request packets on the first port	ntdp timer hop-delay <i>time</i>	Optional 200 ms by default.
Configure the port delay to forward topology collection request	ntdp timer port-delay <i>time</i>	Optional 20 ms by default.

Manually Collecting NTDP Information

The management device collects topology information periodically after a cluster is created. In addition, you can configure to manually collect NTDP information to initiate NTDP information collection, thus managing and monitoring the device on real time, regardless of whether a cluster is created.

Follow these steps to configure to manually collect NTDP information:

To do...	Use the command...	Remarks
Manually collect NTDP information	ntdp explore	Required

Enabling the Cluster Function

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the cluster function globally	cluster enable	Optional Enabled by default.

Establishing a Cluster

Before establishing a cluster, you need to configure a private IP address pool for the devices to be added to the cluster. When a candidate device is added to a cluster, the management device assigns a private IP address to it for the candidate device to communicate with other devices in the cluster. This enables you to manage and maintain member devices in a cluster through the management device.



CAUTION:

- *If the routing table of the management device is full when a cluster is created, that is, entries with the destination address as a candidate device cannot be added to the routing table, all candidate devices will be added to and removed from the cluster repeatedly.*
- *If the routing table of a candidate device is full when the candidate device is added to the cluster, that is, entries with the destination address as the management device cannot be added to the routing table, the candidate device will be added to and removed from the cluster repeatedly.*



CAUTION:

- *You can only specify a management VLAN before establishing a cluster. After a device has been added to the cluster, you cannot modify the management VLAN. To change the management VLAN after the cluster is established, you should remove the cluster on the management device, re-specify the management VLAN and reestablish a cluster.*
- *For the purpose of security, you are not recommended to configure the VLAN ID of the management VLAN as the default VLAN ID of the port connecting the management device to its member devices.*
- *Only when the default VLAN ID of all subtending ports and the port connecting the management device to its member device is that of the management VLAN, can the packets without a tag from the management VLAN pass the ports. Otherwise, you must configure the packets from the management VLAN to pass these ports. For the configuration procedure, refer to “VLAN Configuration” on page 83.*
- *You must configure the IP address pool before establishing a cluster and configure it on the management device only. If a cluster has already been established, you are not allowed to change the IP address pool.*

Manually establishing a cluster

Follow these steps to manually establish a cluster:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Specify the management VLAN	management-vlan <i>vlan-id</i>	Optional By default, VLAN 1 is the management VLAN.
Enter cluster view	cluster	-
Configure the private IP address range for member devices on a device which is to be configured as the management device	ip-pool <i>administrator-ip-address</i> { <i>mask</i> <i>mask-length</i> }	Required For a cluster to work normally, the IP addresses of the VLAN interfaces of the management device and member devices must not be in the same network segment as that of the cluster address pool.
Configure the current device as the management device and assign a name to it	build <i>name</i>	Required By default, the device is not the management device.

Automatically establishing a cluster

In addition to establishing a cluster manually, you are also provided with the means to establish a cluster automatically. With only a few commands (as shown in the table below) on the management device, you can let the system automatically build a cluster.

During the process, you will first be asked to enter a name for the cluster you want to establish, the system then lists all the candidate devices within your predefined hop counts and starts to automatically add them to the cluster.

You can use <Ctrl+C> anytime during the adding process to exit cluster auto-building. However, this will only stop adding new devices into the cluster, and devices already added in the cluster are not removed.

Follow these steps to automatically establish a cluster:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Specify the management VLAN	management-vlan <i>vlan-id</i>	Optional By default, VLAN 1 is the management VLAN.
Enter cluster view	cluster	-
Configure the private IP address range for member devices on a device which is to be configured as the management device	ip-pool <i>administrator-ip-address</i> { <i>mask</i> <i>mask-length</i> }	Required For a cluster to work normally, the IP addresses of the VLAN interfaces of the management device and member devices must not be in the same network segment as the cluster address pool.
Establish a cluster automatically	auto-build [recover]	Required

Configuring Communication Between the Management Device and the Member Devices Within a Cluster

In a cluster, the management device and member devices communicate by sending handshake packets to maintain connection between them. You can configure interval of sending handshake packets and the holdtime of a device on the management device. This configuration applies to all member devices within the cluster.

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the interval to send handshake packets	timer <i>interval-time</i>	Optional 10 seconds by default
Configure the holdtime of a device	holdtime <i>seconds</i>	Optional 60 seconds by default

Configuring the Destination MAC Address of Cluster Management Multicast Packets

By default, the destination MAC address of cluster management multicast packets (including NDP, NTDP and HABP packets) is 010f-e200-0002, which IEEE reserved for later use. Since some devices cannot forward the multicast packets with the destination MAC address of 010f-e200-0002, cluster management packets cannot traverse these devices. For a cluster to work normally in this case, you can modify the destination MAC address of a cluster management multicast packet without changing the current networking.

The management device periodically sends MAC address negotiation broadcast packets to advertise the destination MAC address of the cluster management multicast packets.

Follow these steps to configure the destination MAC address of the cluster management multicast packets:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter cluster view	cluster	-
Configure the destination MAC address for cluster management multicast packets	cluster-mac <i>mac-address</i>	Required The destination MAC address is 010f-e200-0002 by default.
Configure the interval to send MAC address negotiation broadcast packets for cluster management multicast packets	cluster-mac syn-interval <i>interval-time</i>	Optional One minute by default.

Configuring Cluster Member Management

Adding/Removing a member device

You can manually add a candidate device to a cluster, or remove a member device from a cluster. These operations must be done through the management device, otherwise you will be prompted with an error message.

Follow these steps to add/remove a member device:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Enter cluster view	cluster	-
Add a candidate device to the cluster	add-member [<i>member-number</i>] mac-address <i>mac-address</i> [password <i>password</i>]	Optional
Remove a member device from the cluster	delete-member <i>member-number</i> [to-black-list]	Required

Rebooting a member device

Communication between the management and member devices may be interrupted due to some configuration errors. Through the remote control function of member devices, you can control them remotely on the management device. For example, you can reboot a member device that operates improperly and specify to delete the booting configuration file when the member device reboots, and thus achieve normal communication between the management and member devices.

Follow these steps to reboot a member device:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter cluster view	cluster	-
Reboot a specified member device	reboot member { <i>member-number</i> } mac-address <i>mac-address</i> } [eraseflash]	Required

Configuring the Member Devices

Enabling NDP Globally and for Specific Ports

Refer to “Enabling NDP Globally and for Specific Ports” on page 912.

Enabling NTDP Globally and for Specific Ports

Refer to “Enabling NTDP Globally and for Specific Ports” on page 912.

Manually Collecting NTDP Information

Refer to “Manually Collecting NTDP Information” on page 913.

Enabling the Cluster Function

Refer to “Enabling the Cluster Function” on page 914.

Deleting a Member Device from a Cluster

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter cluster view	cluster	-

To do...	Use the command...	Remarks
Delete a member device from the cluster	undo administrator-address	Required

Configuring Access Between the Management Device and Its Member Devices

After having successfully configured NDP, NTDP and cluster, you can configure, manage and monitor the member devices through the management device. You can manage member devices in a cluster through switching from the operation interface of the management device to that of a member device or configure the management device by switching from the operation interface of a member device to that of the management device.

Follow these steps to configure access between member devices of a cluster:

To do...	Use the command...	Remarks
Switch from the operation device of the management device to that of a member device	cluster switch-to { <i>member-number</i> mac-address <i>mac-address</i> sysname <i>member-sysname</i> }	Required
Switch from the operation interface of a member device to that of the management device	cluster switch-to administrator	Required



CAUTION: *Telnet connection is used on the switch between the management device and member devices. Note the following when switching between them*

- *Before the switch, execute the **telnet server enable** command to enable Telnet. Otherwise, the switch fails.*
- *Authentication is required when you switch a member device to the management device. The switch fails if authentication is not passed. Your user level is allocated according to the predefined level by the management device if authentication is passed.*
- *When a candidate device is added to a cluster and becomes a member device, its super password will be automatically synchronized to the management device. Therefore, after a cluster is established, you are not recommended to modify the super password of the member device (including management device and member devices) of the cluster; otherwise, the switch may fail because of authentication failure.*
- *When you switch the management device to a member device, if member *n* does not exist, the system prompts error; if the switch succeeds, your user level on the management device is retained.*
- *If the Telnet users on the device to be logged in reach the maximum number, the switch fails.*
- *To prevent resource waste, avoid recycling switch when configuring access between cluster members. For example, if you switch from the operation interface of the management device to that of a member device and then need to switch back to that of the management device, use the **quit** command to end the switch, but not the **cluster switch-to administrator** command to switch to the operation interface of the management device.*

Adding a Candidate Device to a Cluster

Follow these steps to add a candidate device to a cluster:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter cluster view	cluster	-
Add a candidate device to the cluster	administrator-address <i>mac-address name name</i>	Required

Configuring Advanced Cluster Functions

This section covers these topics:

- “Configuring Topology Management” on page 919
- “Configuring Interaction for a Cluster” on page 920

Configuring Topology Management

The concepts of blacklist and whitelist are used for topology management. An administrator can diagnose the network by comparing the current topology and the standard topology.

- Current topology: The information of a node and its neighbors of the cluster.
- Topology management whitelist (standard topology): A whitelist is a list of topology information that has been confirmed by the administrator as correct. You can get the information of a node and its neighbors from the current topology. Based on the information, you can manage and maintain the whitelist by adding, deleting or modifying a node.
- Topology management blacklist: A blacklist is a list of devices that are not allowed to join a cluster unless the administrator manually removes them from the list. A blacklist contains the MAC addresses of devices. If a blacklist device is connected to network through another device not included in the blacklist, the MAC address and access port of the latter are also included in the blacklist.

A whitelist member cannot be a blacklist member, and vice versa. However, a topology node can belong to neither the whitelist nor the blacklist. Nodes of this type are usually newly added nodes, whose identities are to be confirmed by the administrator.

You can back up the whitelist and blacklist to prevent them from missing when a power failure occurs to the management device. The following two backup and restore mechanisms are available:

- Backing them up on the FTP server shared by the cluster. You can manually restore the whitelist and blacklist from the FTP server.
- Backing them up in the Flash of the management device. When the management device restarts, the whitelist and blacklist will be automatically restored from the Flash. When a cluster is reestablished, you can choose whether to restore the whitelist and blacklist from the Flash automatically, or you can manually restore them from the Flash of the management device.

Follow these steps to configure cluster topology management:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter cluster view	cluster	-
Add a device to the blacklist	black-list add-mac <i>mac-address</i>	Optional
Remove a device from the blacklist	black-list delete-mac { all <i>mac-address</i> }	Optional
Confirm the current topology and save it as the standard topology	topology accept { all [save-to { ftp-server local-flash }] mac-address <i>mac-address</i> member-id <i>member-number</i> }	Optional
Save the standard topology to the FTP server or the local Flash	topology save-to { ftp-server local-flash }	Optional
Restore the standard topology information from the FTP server or the local Flash	topology restore-from { ftp-server local-flash }	Optional You must ensure that the topology is correct before restoring it as the device itself cannot judge the correctness in topology.

Configuring Interaction for a Cluster

After establishing a cluster, you can configure FTP/TFTP server, NM host and log host for the cluster on the management device.

- After you configure an FTP/TFTP server for a cluster, the members in the cluster access the FTP/TFTP server configured through the management device.
- After you configure a log host for a cluster, all the log information of the members in the cluster will be output to the configured log host in the following way: first, the member devices send their log information to the management device, which then converts the addresses of log information and sends them to the log host.
- After you configure an NM host for a cluster, the member devices in the cluster send their Trap messages to the shared SNMP NM host through the management device.

If the port of an access NM device (including FTP/TFTP server, NM host and log host) does not allow the packets from the management VLAN to pass, the NM device cannot manage the devices in a cluster through the management device. In this case, on the management device, you need to configure the VLAN interface of the access NM device (including FTP/TFTP server, NM host and log host) as the NM interface.

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter cluster view	cluster	-
Configure the FTP server shared by the cluster by setting an IP address, username and password	ftp-server <i>ip-address</i> [user-name <i>username</i> password { simple cipher } <i>password</i>]	Required By default, no FTP server is configured for a cluster.

To do...	Use the command...	Remarks
Configure the TFTP server shared by the member devices in the cluster	tftp-server <i>ip-address</i>	Required By default, no TFTP server is configured for a cluster.
Configure the log host shared by the member devices in the cluster	logging-host <i>ip-address</i>	Required By default, no log host is configured for a cluster.
Configure the SNMP NM host shared by the member devices in the cluster	snmp-host <i>ip-address</i> [community-string read <i>string1</i> write <i>string2</i>]	Required By default, no SNMP host is configured.
Configure the NM interface of the management device	nm-interface vlan-interface <i>vlan-interface-id</i>	Optional

**CAUTION:**

- For the configured log host to take effect, you must execute the **info-center loghost** command in system view first. For more information about the **info-center loghost** command, refer to "Configuring Information Center" on page 1009.
- To isolate management protocol packets of a cluster from packets outside the cluster, you are recommended to configure to prohibit packets from the management VLAN from passing the ports that connect the management device with the devices outside the cluster and configure the NM interface for the management device.

Displaying and Maintaining Cluster Management

To do...	Use the command...	Remarks
Display NDP configuration information	display ndp [interface <i>interface-list</i>]	Available in any view
Display the global NTDP information	display ntdp	
Display the device information collected through NTDP	display ntdp device-list [verbose]	
Display the detailed NTDP information of a specified device	display ntdp single-device mac-address <i>mac-address</i>	
View cluster state and statistics	display cluster	
View the standard topology information	display cluster base-topology [mac-address <i>mac-address</i> member-id <i>member-number</i>]	
View the current blacklist of the cluster	display cluster black-list	
View the information of candidate devices	display cluster candidates [mac-address <i>mac-address</i> verbose]	
Display the current topology information or the topology path between two devices	display cluster current-topology [mac-address <i>mac-address</i> to-mac-address <i>mac-address</i>] member-id <i>member-number</i> to-member-id <i>member-number</i>]]	
Display members in a cluster	display cluster members [<i>member-number</i> verbose]	
Clear NDP statistics	reset ndp statistics [interface <i>interface-list</i>]	Available in user view



- Support for the **display ntdp single-device mac-address** command varies with devices.
- When you display the cluster topology information, the devices attached to the switch that is listed in the blacklist will not be displayed.

Cluster Management Configuration Examples

Cluster Management Configuration Example One

Network requirements

Three switches form a cluster, in which:

- One device serves as the management device.
- The other two are the member devices.

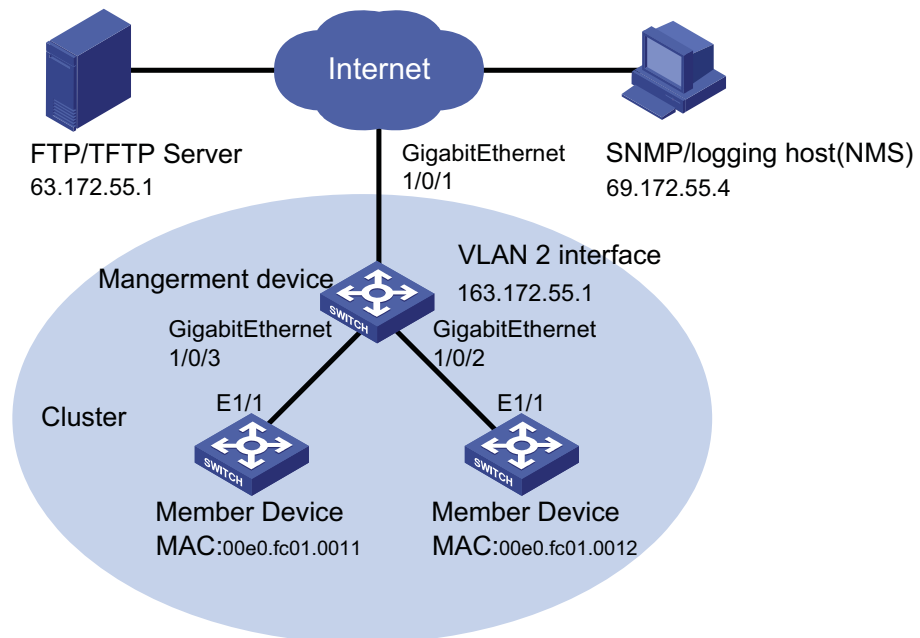
The specific requirements are as follows:

- The management device is connected to the external network through its Ethernet 1/1 port. The two member devices are connected to Ethernet 1/2 and Ethernet 1/3 ports of the management device.

- Ethernet 1/1 port of the management device belongs to VLAN 2, whose interface IP address is 163.172.55.1/24. The network management interface of the management device is VLAN-interface 2. VLAN 2 is the network management (NM) interface of the management device.
- All the devices in the cluster use the same FTP server and TFTP server, which share one IP address: 63.172.55.1/24.
- The SNMP NMS and log host share one IP address: 69.172.55.4/24.
- The management VLAN of the cluster is VLAN 10.
- Add the device whose MAC address is 00E0-FC01-0013 to the blacklist.

Network diagram

Figure 270 Network diagram for cluster management



Configuration procedure

- 1 Configuring the member device (All member devices have the same configuration, taking one member as an example)

Enable NDP globally and for the Ethernet1/1 port.

```
<Switch> system-view
[Switch] ndp enable
[Switch] interface Ethernet1/1
[Switch- Ethernet1/1] ndp enable
[Switch- Ethernet1/1] quit
```

Enable NTDP globally and for the Ethernet1/0/1 port.

```
[Switch] ntdp enable
[Switch] interface Ethernet1/1
[Switch-Ethernet1/1] ntdp enable
[Switch-Ethernet1/1] quit
```

Enable the cluster function.

```
[Switch] cluster enable
```

2 Configuring the management device

Enable NDP globally and for the GigabitEthernet1/0/2,GigabitEthernet1/0/3 ports.

```
<Switch> system-view
[Switch] ndp enable
[Switch] interface GigabitEthernet1/0/2
[Switch-GigabitEthernet1/0/2] ndp enable
[Switch-GigabitEthernet1/0/2] quit
[Switch] interface GigabitEthernet1/0/3
[Switch-GigabitEthernet1/0/3] ndp enable
[Switch-GigabitEthernet1/0/3] quit
```

Configure the period for the receiving device to keep NDP packets as 200 seconds.

```
[Switch] ndp timer aging 200
```

Configure the interval to send NDP packets as 70 seconds.

```
[Switch] ndp timer hello 70
```

Enable NTDP globally and for the GigabitEthernet1/0/2,GigabitEthernet1/0/3 ports.

```
[Switch] ntdp enable
[Switch] interface GigabitEthernet1/0/2
[Switch-GigabitEthernet1/0/2] ntdp enable
[Switch-GigabitEthernet1/0/2] quit
[Switch] interface GigabitEthernet1/0/3
[Switch-GigabitEthernet1/0/3] ntdp enable
[Switch-GigabitEthernet1/0/3] quit
```

Configure the hop count to collect topology as 2.

```
[Switch] ntdp hop 2
```

Configure the delay time for topology-collection request packets to be forwarded on member devices as 150 ms.

```
[Switch] ntdp timer hop-delay 150
```

Configure the delay time for topology-collection request packets to be forwarded through the ports of member devices as 15 ms.

```
[Switch] ntdp timer port-delay 15
```

Configure the interval to collect topology information as 3 minutes.

```
[Switch] ntdp timer 3
```

Configure the management VLAN of the cluster as VLAN 10.

```
[Switch] vlan 10
[Switch-vlan10] quit
[Switch] management-vlan 10
```

Configure the port connecting the management device to candidate devices as a Trunk port and allow packets from the management VLAN to pass.

```
[Switch] interface GigabitEthernet 1/0/2
[Switch-GigabitEthernet 1/0/2] port link-type trunk
[Switch-GigabitEthernet 1/0/2] port trunk permit vlan 10
[Switch-GigabitEthernet 1/0/2] quit
[Switch] interface GigabitEthernet 1/0/3
[Switch-GigabitEthernet 1/0/3] port link-type trunk
```



```

[Switch-GigabitEthernet 1/0/3] port trunk permit vlan 10
[Switch-GigabitEthernet 1/0/3] quit
# Enable the cluster function.
[Switch] cluster enable
# Enter cluster view.
[Switch] cluster
# Configure an IP address pool for the cluster. The IP address pool contains six IP
addresses, starting from 172.16.0.1.
[Switch-cluster] ip-pool 172.16.0.1 255.255.255.248
# Specify a name for the cluster and create the cluster.
[Switch-cluster] build aabbcc
Restore topology from local flash file, for there is no base topology
.
(Please confirm in 30 seconds, default No). (Y/N)
N
# Enable management VLAN auto-negotiation.
[aabbcc_0.Switch-cluster] management-vlan synchronization enable
# Configure the holdtime of the member device information as 100 seconds.
[aabbcc_0.Switch-cluster] holdtime 100
# Configure the interval to send handshake packets as 10 seconds.
[aabbcc_0.Switch-cluster] timer 10
# Configure the FTP Server, TFTP Server, Log host and SNMP host for the cluster.
[aabbcc_0.Switch-cluster] ftp-server 63.172.55.1
[aabbcc_0.Switch-cluster] tftp-server 63.172.55.1
[aabbcc_0.Switch-cluster] logging-host 69.172.55.4
[aabbcc_0.Switch-cluster] snmp-host 69.172.55.4
# Add the device whose MAC address is 00E0-FC01-0013 to the blacklist.
[aabbcc_0.Switch-cluster] black-list add-mac 00e0-fc01-0013
[aabbcc_0.Switch-cluster] quit
# Configure the network management interface.
[aabbcc_0.Switch] vlan 2
[aabbcc_0.Switch-vlan2] port GigabitEthernet 1/0/1
[aabbcc_0.Switch] quit
[aabbcc_0.Switch] interface vlan-interface 2
[aabbcc_0.Switch-Vlan-interface2] ip address 163.172.55.1 24
[aabbcc_0.Switch-Vlan-interface2] quit
[aabbcc_0.Switch] cluster
[aabbcc_0.Switch-cluster] nm-interface vlan-interface 2

```



- Upon completion of the above configurations, you can execute the **cluster switch-to** { member-number | **mac-address** mac-address } command on the management device to switch to the operation interface of a member device to maintain and manage it. You can then execute the **quit** command to return to the operation interface of the management device.
- You can also reboot a member device by executing the **reboot member** command on the management device.

- You can execute the **cluster switch-to administrator** command to switch to the operation interface of the management device.
- For detailed information about these configurations, refer to the preceding description in this chapter.

74

UDP HELPER CONFIGURATION

When configuring UDP Helper, go to these sections for information you are interested in:

- "Introduction to UDP Helper" on page 927
- "Configuring UDP Helper" on page 927
- "Displaying and Maintaining UDP Helper" on page 928
- "UDP Helper Configuration Example" on page 928

Introduction to UDP Helper

Sometimes, a host needs to forward broadcasts to obtain network configuration information or request the names of other devices on the network. However, if the server or the device to be requested is located in another broadcast domain, the host cannot obtain such information through broadcast.

To solve this problem, the device provides the UDP Helper function to relay specified UDP packets. In other words, UDP Helper functions as a relay agent that converts UDP broadcast packets into unicast packets and forwards them to a specified destination server.

With UDP Helper enabled, the device decides whether to forward a received UDP broadcast packet according to the UDP destination port number of the packet.

- If the destination port number of the packet matches the one pre-configured on the device, the device modifies the destination IP address in the IP header, and then sends the packet to the specified destination server.
- If not, the device sends the packet to the upper layer protocol for processing.

Configuring UDP Helper

Follow these steps to configure UDP Helper:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable UDP Helper	udp-helper enable	Required Disabled by default.
Enable the forwarding of packets with the specified UDP destination port number(s)	udp-helper port { <i>port-number</i> dns netbios-ds netbios-ns tacacs tftp time }	Required By default, no UDP port number is specified.
Enter VLAN interface view	interface Vlan-interface <i>vlan-id</i>	-

To do...	Use the command...	Remarks
Specify the destination server to which UDP packets are to be forwarded	udp-helper server <i>ip-address</i>	Required No destination server is specified by default.



CAUTION:

- The UDP Helper enabled device cannot forward DHCP broadcast packets. That is to say, the UDP port number cannot be set to 67 or 68.
- For the **dns**, **netbios-ds**, **netbios-ns**, **tacacs**, **tftp**, and **time** keywords, you can specify port numbers or the corresponding parameters. For example, **udp-helper port 53** and **udp-helper port dns** specify the same UDP port number.
- The configuration of all UDP ports is removed if you disable UDP Helper.
- You can configure up to 20 destination servers on a VLAN interface.

Displaying and Maintaining UDP Helper

To do...	Use the command...	Remarks
Displays the information of forwarded UDP packets	display udp-helper server [interface <i>Vlan-interface</i> <i>vlan-id</i>]	Available in any view
Clear statistics about packets forwarded	reset udp-helper packet	Available in user view

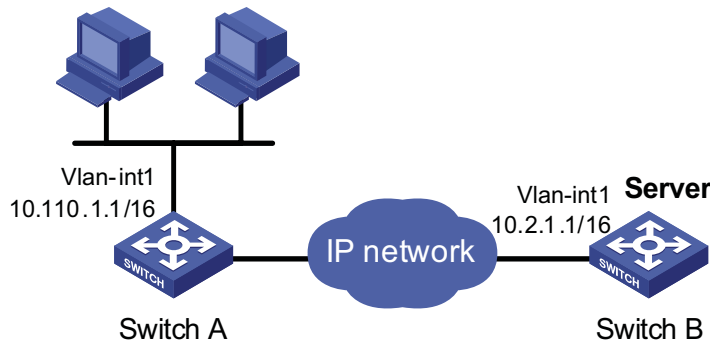
UDP Helper Configuration Example

Network requirements

The interface VLAN-interface 1 of Switch A has the IP address of 10.110.1.1/16, connecting to the network segment 10.110.0.0/16. Enable the forwarding of broadcast packets with the UDP destination port number 55 to the destination server 10.2.1.1/16.

Network diagram

Figure 271 Network diagram for UDP Helper configuration



Configuration procedure



The following configuration assumes that a route from Switch A to the network segment 10.2.0.0/16 is available.

Enable UDP Helper.

```
<SwitchA> system-view  
[SwitchA] udp-helper enable
```

Enable the forwarding broadcast packets with the UDP destination port number 55.

```
[SwitchA] udp-helper port 55
```

Specify the server with the IP address of 10.2.1.1 as the destination server to which UDP packets are to be forwarded.

```
[SwitchA] interface vlan-interface 1  
[SwitchA-Vlan-interface1] ip address 10.110.1.1 16  
[SwitchA-Vlan-interface1] udp-helper server 10.2.1.1
```


When configuring SNMP, go to these sections for information you are interested in:

- “SNMP Overview” on page 931
- “SNMP Configuration” on page 933
- “Configuring SNMP Logging” on page 935
- “Trap Configuration” on page 936
- “Displaying and Maintaining SNMP” on page 937
- “SNMP Configuration Example” on page 938
- “SNMP Logging Configuration Example” on page 939

SNMP Overview

Simple Network Management Protocol (SNMP) offers a framework to monitor network devices through TCP/IP protocol suite. It provides a set of basic operations in monitoring and maintaining the Internet and has the following characteristics:

- Automatic network management: SNMP enables network administrators to search and modify information, find and diagnose network problems, plan for network growth, and generate reports on network nodes.
- SNMP shields the physical differences between various devices and thus realizes automatic management of products from different manufacturers. Offering only the basic set of functions, SNMP makes the management tasks independent of both the physical features of the managed devices and the underlying networking technology. Thus, SNMP achieves effective management of devices from different manufacturers, especially in small, high-speed and low cost network environments.

SNMP Mechanism

An SNMP enabled network comprises network management station (NMS) and Agent.

- NMS is a station that runs the SNMP client software. It offers a user friendly interface, making it easier for network administrators to perform most network management tasks. Currently, the most commonly used NMSs include Quidview, Sun NetManager, and IBM NetView.
- Agent is a program on the device. It receives and handles requests sent from the NMS. Only under certain circumstances, such as interface state change, will the Agent inform the NMS.
- NMS manages an SNMP enabled network, whereas Agent is the managed network device. They exchange management information through the SNMP protocol.

SNMP provides the following four basic operations:

- Get operation: NMS gets the value of a certain variable of Agent through this operation.
- Set operation: NMS can reconfigure certain values in the Agent MIB (Management Information Base) to make the Agent perform certain tasks by means of this operation.
- Trap operation: Agent sends Traps to the NMS through this operation.
- Inform operation: NMS sends Traps to other NMSs through this operation.

SNMP Protocol Version

Currently, SNMP agents support SNMPv3 and are compatible with SNMPv1 and SNMPv2c.

- SNMPv1 authenticates by means of community name, which defines the relationship between an SNMP NMS and an SNMP Agent. SNMP packets with community names that did not pass the authentication on the device will simply be discarded. A community name performs a similar role as a key word and can be used to regulate access from NMS to Agent.
- SNMPv2c authenticates by means of community name. Compatible with SNMPv1, it extends the functions of SNMPv1. SNMPv2c provides more operation modes such as GetBulk and InformRequest; it supports more data types such as Counter64 and Counter32; and it provides various error codes, thus being able to distinguish errors in more detail.
- SNMPv3 offers an authentication that is implemented with a User-Based Security Model (USM). You can set the authentication and privacy functions. The former is used to authenticate the validity of the sending end of the authentication packets, preventing access of illegal users; the latter is used to encrypt packets between the NMS and Agent, preventing the packets from being intercepted. USM ensures a more secure communication between SNMP NMS and SNMP Agent by authentication with privacy, authentication without privacy, or no authentication no privacy.

Successful interaction between NMS and Agent requires consistency of SNMP versions configured on them. You can configure multiple SNMP versions for an Agent to interact with different NMSs.

MIB Overview

Any managed resource can be identified as an object, which is known as the managed object. Management Information Base (MIB) is a collection of all the managed objects. It defines a set of characteristics associated with the managed objects, such as the object identifier (OID), access right and data type of the objects. Each Agent has its own MIB. NMS can read or write the managed objects in the MIB. The relationship between NMS, Agent and MIB is shown in Figure 272.

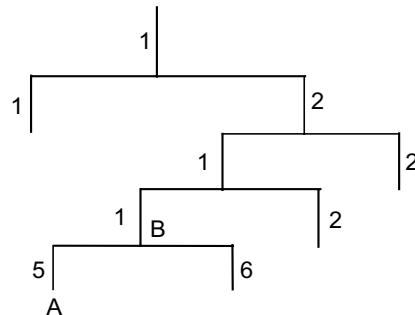
Figure 272 Relationship between NMS, Agent and MIB



MIB stores data using a tree structure. The node of the tree is the managed object and can be uniquely identified by a path starting from the root node. As illustrated in the following figure, the managed object B can be uniquely identified by a

string of numbers {1.2.1.1}. This string of numbers is the OID of the managed object B.

Figure 273 MIB tree



SNMP Configuration

As configurations for SNMPv3 differ substantially from those of SNMPv1 and SNMPv2c, their SNMP functionalities will be introduced separately below.

Follow these steps to configure SNMPv3:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable SNMP Agent	snmp-agent	Optional Disabled by default You can enable SNMP Agent through this command or any commands that begin with snmp-agent .
Configure SNMP Agent system information	snmp-agent sys-info { contact <i>sys-contact</i> location <i>sys-location</i> version { all { v1 v2c v3 }* }	Optional The defaults are as follows: 3Com Corporation for contact, Marlborough, MA for location, and SNMP v3 for the version.
Configure an SNMP agent group	snmp-agent group v3 <i>group-name</i> [authentication privacy] [read-view <i>read-view</i>] [write-view <i>write-view</i>] [notify-view <i>notify-view</i>] [acl <i>acl-number</i>]	Required
Convert the user-defined plain text password to a cipher text password	snmp-agent calculate-password <i>plain-password</i> mode { md5 sha } { local-switch fabricid specified-switch fabricid <i>string</i> }	Optional

To do...	Use the command...	Remarks
Add a new user to an SNMP agent group	snmp-agent usm-user v3 <i>user-name group-name</i> [[cipher] authentication-mode { md5 sha } <i>auth-password</i> [privacy-mode { aes128 des56 } <i>priv-password</i>]] [acl <i>acl-number</i>]	Required If the cipher keyword is specified, the arguments <i>auth-password</i> and <i>priv-password</i> are considered as cipher text password.
Configure the maximum size of an SNMP packet that can be received or sent by an SNMP agent	snmp-agent packet max-size <i>byte-count</i>	Optional 1,500 bytes by default
Configure the switch fabric ID for a local SNMP agent	snmp-agent local-switch fabricid <i>switch fabricid</i>	Optional Company ID and device ID by default
Create or update the MIB view content for an SNMP agent	snmp-agent mib-view { excluded included } <i>view-name oid-tree</i> [mask <i>mask-value</i>]	Optional MIB view name is ViewDefault and OID is 1 by default.

Follow these steps to configure SNMPv1 and SNMPv2c:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable SNMP Agent	snmp-agent	Required Disabled by default You can enable SNMP Agent through this command or any commands that begin with snmp-agent .
Configure SNMP Agent system information	snmp-agent sys-info { contact <i>sys-contact</i> location <i>sys-location</i> version { { v1 v2c v3 }* all } }	Required The defaults are as follows: 3Com Corporation for contact, Marlborough, MA for location and SNMP v3 for the version.
Configure SNMP NMS access right	snmp-agent community { read write } <i>community-name</i> [acl <i>acl-number</i> mib-view <i>view-name</i>]*	Use either approach. Both commands can be used to configure SNMP NMS access rights. The second command was introduced to be compatible with SNMPv3.
Configure an indirect group	snmp-agent group { v1 v2c } <i>group-name</i> [read-view <i>read-view</i>] [write-view <i>write-view</i>] [notify-view <i>notify-view</i>] [acl <i>acl-number</i>]	The community name configured on NMS should be consistent with the username configured on the Agent.
Add a new user to an SNMP group	snmp-agent usm-user { v1 v2c } <i>user-name group-name</i> [acl <i>acl-number</i>]	

To do...	Use the command...	Remarks
Configure the maximum size of an SNMP packet that can be received or sent by an SNMP agent	snmp-agent packet max-size <i>byte-count</i>	Optional 15,00 bytes by default
Configure the switch fabric ID for a local SNMP agent	snmp-agent local-switch fabricid <i>switch fabricid</i>	Optional Company ID and device ID by default
Create or update MIB view content for an SNMP agent	snmp-agent mib-view { excluded included } <i>view-name oid-tree</i> [mask <i>mask-value</i>]	Optional ViewDefault by default



CAUTION: The validity of a USM user depends on the switch fabric ID of the SNMP agent. If the switch fabric ID used for USM user creation is not identical to the current switch fabric ID, the USM user is invalid.

Configuring SNMP Logging

Introduction to SNMP Logging

SNMP logs the GET and SET operations that NMS performs to SNMP Agent. When the GET operation is performed, Agent logs the IP address of NMS, node name of the GET operation and OID of the node. When the SET operation is performed, Agent logs the IP address of NMS, node name of the SET operation, OID of the node, the value set and the error code and index returned with the SET operation. These logs will be transferred to system information and sent to the information center to be checked and tracked.

SNMP logs GET request, SET request and SET response, but does not log GET response.

Enabling SNMP Logging

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable SNMP logging	snmp-agent log { all get-operation set-operation }	Required Disabled by default.
Configure SNMP log output rules	info-center source { <i>module-name</i> default } channel { <i>channel-number</i> <i>channel-name</i> } [debug { level <i>severity</i> state <i>state</i> } * log { level <i>severity</i> state <i>state</i> } * trap { level <i>severity</i> state <i>state</i> } *] *	Optional By default, SNMP logs are output to loghost and logfile only. To output SNMP logs to other destinations such as console or monitor terminal, you need to set the output destinations with this command.



- Logs occupy storage space of the device, thus affecting the performance of the device. Therefore, you are recommended to disable SNMP logging.
- The priority of SNMP log is informational, meaning it is a common prompt of the device. To check SNMP logs, enable the information center to output system information with the severity of informational.

- The size of SNMP logs cannot exceed that allowed by the information center and the sum of the node, and value field of each log information cannot exceed 1K bytes; otherwise, the exceeded part will be output.
- For the detailed description of system information, the information center and the **info-center source** command, refer to “Configuring Information Center” on page 1009.

Trap Configuration

SNMP Agent sends Traps to NMS to alert the latter of critical and important events (such as restart of the managed device).

Configuration Prerequisites

Basic SNMP configurations have been completed. These configurations include version configuration: community name is needed when SNMPv1 and v2c are adopted; username and MIB view are needed if SNMPv3 is adopted.

Configuration Procedure

Enabling Trap transmission

Follow these steps to enable Trap transmission:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Set to enable the device to send Traps globally	snmp-agent trap enable [bgp configuration flash ospf [<i>process-id</i>] standard [authentication coldstart linkdown linkup warmstart]* system voice vrrp [authfailure newmaster]]	Optional All types of Traps are allowed by default.
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Set to enable the device to send Traps of interface state change	enable snmp trap updown	Optional Transmission of Traps of interface state change is allowed by default.



CAUTION: To enable an interface to send SNMP Traps when its state changes, you need to enable the Link up/down Trap packet transmission function on an interface and globally. Use the **enable snmp trap updown** command to enable this function on an interface, and use the **snmp-agent trap enable [standard [linkdown | linkup] *]** command to enable this function globally.

Configuring Trap transmission parameters

Follow these steps to configure Trap:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Configure target host attribute for Traps	snmp-agent target-host trap address udp-domain { ip-address ipv6 ipv6-address } [udp-port port-number] params securityname security-string [v1 v2c v3 [authentication privacy]]	Required
Configure the source address for Traps	snmp-agent trap source interface-type { interface-number interface-number.subnumber }	Optional
Extend the standard linkUp/linkDown Traps defined in RFC	snmp-agent trap if-mib link extended	Optional Standard linkUp/linkDown Traps defined in RFC are used by default.
Configure the queue size for sending Traps	snmp-agent trap queue-size size	Optional 100 by default
Configure the lifetime for Traps	snmp-agent trap life seconds	Optional 120 seconds by default



The extended linkUp/linkDown Traps comprise the standard linkUp/linkDown Traps defined in RFC plus interface description and interface type. If the extended messages are not supported on NMS, you can disable this function and enable the device to send standard linkUp/linkDown Traps.

Displaying and Maintaining SNMP

To do...	Use the command...	Remarks
Display SNMP-agent system information, including the contact, location, and version of the SNMP	display snmp-agent sys-info [contact location version]*	Available in any view
Display SNMP agent statistics	display snmp-agent statistics	
Display the SNMP agent switch fabric ID	display snmp-agent local-switch fabricid	
Display SNMP agent group information	display snmp-agent group [group-name]	
Display the modules that can send Traps and whether their Trap sending is enabled or not	display snmp-agent trap-list	
Display SNMP v3 agent user information	display snmp-agent usm-user [switch fabricid switch fabricid username user-name group group-name] *	
Display SNMP v1 or v2c agent community information	display snmp-agent community [read write]	
Display MIB view information for an SNMP agent	display snmp-agent mib-view [exclude include viewname view-name]	

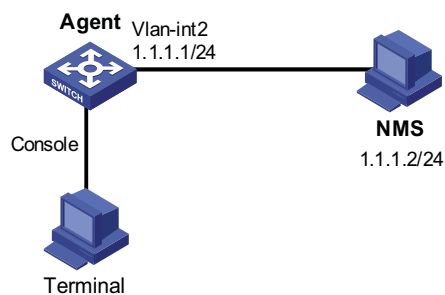
SNMP Configuration Example

Network requirements

- The NMS connects to the agent, a switch, through an Ethernet.
- The IP address of the NMS is 1.1.1.2/24.
- The IP address of VLAN interface on the switch is 1.1.1.1/24.
- NMS monitors and manages Agent using SNMPv2c. Agent reports errors or faults to the NMS.

Network diagram

Figure 274 Network diagram for SNMP (on a switch)



Configuration procedure

1 Configuring SNMP Agent

Configure the SNMP basic information, including version and community name.

```

<Sysname> system-view
[Sysname] snmp-agent sys-info version v2c
[Sysname] snmp-agent community read public
[Sysname] snmp-agent community write private
  
```

Configure VLAN-interface 2 (with the IP address of 1.1.1.1/24). Add the port Ethernet 1/0 to VLAN 2.

```

[Sysname] vlan 2
[Sysname-vlan2] port ethernet 1/0
[Sysname-vlan2] interface vlan-interface 2
[Sysname-Vlan-interface2] ip address 1.1.1.1 255.255.255.0
[Sysname-Vlan-interface2] quit
  
```

Configure the contact person and physical location information of the switch.

```

[Sysname] snmp-agent sys-info contact Mr.Wang-Tel:3306
[Sysname] snmp-agent sys-info location telephone-closet,3rd-floor
  
```

Enable the sending of Traps to the NMS with an IP address of 1.1.1.2/24, using **public** as the community name.

```

[Sysname] snmp-agent trap enable
[Sysname] snmp-agent target-host trap address udp-domain 1.1.1.2 udp-port 5000 params securityname public
  
```

2 Configuring SNMP NMS

With SNMPv2c, the user needs to specify the read only community, the read and write community, the timeout time, and number of retries. The user can inquire and configure the device through the NMS.



The configurations on the agent and the NMS must match.

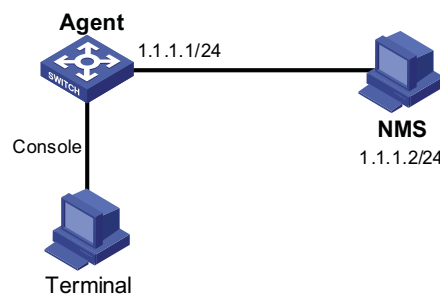
SNMP Logging Configuration Example

Network requirements

- NMS and Agent are connected through an Ethernet
- The IP address of NMS is 1.1.1.2/24
- The IP address of the VLAN interface on Agent is 1.1.1.1/24
- Configure community name, access right and SNMP version on Agent

Network diagram

Figure 275 Network diagram for SNMP logging



Configuration procedure



The configurations for NMS and Agent are omitted.

Enable logging display on the terminal (optional, enabled by default).

```
<Sysname> terminal monitor
<Sysname> terminal logging
```

Enable the information center to output the system information with the severity of informational to the Console port.

```
<Sysname> system-view
[Sysname] info-center source snmp channel console log level informational
```

Enable SNMP logging on Agent to log the GET and SET operations of NMS.

```
[Sysname] snmp-agent log get-operation
[Sysname] snmp-agent log set-operation
```

- The following log information is displayed on the terminal when NMS performs the GET operation to Agent.

```
%Jan 1 02:49:40:566 2006 Sysname SNMP/6/GET:
seqNO = <10> srcIP = <1.1.1.2> op = <get> node = <sysName(1.3.6.1.2.1.1.5.0)> value=<>
```

- The following log information is displayed on the terminal when NMS performs the SET operation to Agent.

```
%Jan 1 02:59:42:576 2006 Sysname SNMP/6/SET:
seqNO = <11> srcIP = <1.1.1.2> op = <set> errorIndex = <0> errorStatus =<noError> node = <sysName(1.3.6.1.2.1.1.5.0)> value = <Sysname>
```

Table 74 Descriptions on the output field of SNMP log

Field	Description
Jan 1 02:49:40:566 2006	The time when SNMP log is generated
seqNO	Sequence number of the SNMP log ()
srcIP	IP address of NMS
op	SNMP operation type (GET or SET)
node	Node name of the SNMP operations and OID of the instance
erroIndex	Error index, with 0 meaning no error
errorstatus	Error status, with noError meaning no error
value	Value set when the SET operation is performed (This field is initialized, meaning the value obtained with the GET operation is not logged.) When the value is a string of characters and the string contains characters not in the range of ASCII 0 to 127 or invisible characters, the string is displayed in hexadecimal. For example, value = <81-43>[hex]



The system information of the information center can be output to the terminal or to the log buffer. In this example, SNMP log is output to the terminal. To set the SNMP log to be output to other directions, refer to "Configuring Information Center" on page 1009.

76

RMON CONFIGURATION

When configuring RMON, go to these sections for information you are interested in:

- "RMON Overview" on page 941
- "Configuring RMON" on page 943
- "Displaying and Maintaining RMON" on page 944
- "RMON Configuration Example" on page 945

RMON Overview

This section covers these topics:

- "Introduction" on page 941
- "RMON Groups" on page 942

Introduction

Remote Monitoring (RMON) is a type of IETF-defined MIB. It is the most important enhancement to the MIB II standard. It allows you to monitor traffic on network segments and even the entire network.

RMON is implemented based on the Simple Network Management Protocol (SNMP) and is fully compatible with the existing SNMP framework.

RMON provides an efficient means of monitoring subnets and allows SNMP to monitor remote network devices in a more proactive and effective way. It reduces traffic between network management station (NMS) and agent, facilitating large network management.

RMON comprises two parts: NMSs and agents running on network devices.

- Each RMON NMS administers the agents within its administrative domain.
- An RMON agent resides on a network monitor or probe for an interface. It monitors and gathers information about traffic over the network segment connected to the interface to provide statistics about packets over a specified period and good packets sent to a host for example.

Working Mechanism

RMON allows multiple monitors. A monitor provides two ways of data gathering:

- Using RMON probes. NMSs can obtain management information from RMON probes directly and control network resources. In this approach, RMON NMSs can obtain all RMON MIB information.
- Embedding RMON agents in network devices such as routers, switches, and hubs to provide the RMON probe function. RMON NMSs exchange data with RMON agents with basic SNMP commands to gather network management information, which, due to system resources limitation, may not cover all MIB

information but four groups of information, alarm, event, history, and statistics, in most cases.

The device adopts the second way. By using RMON agents on network monitors, an NMS can obtain information about traffic size, error statistics, and performance statistics for network management.

RMON Groups Among the ten RMON groups defined by RMON specifications (RFC 1757), 3Com series Ethernet switches support the event group, alarm group, history group and statistics group. Besides, 3Com also defines and implements the private alarm group, which enhances the functions of the alarm group. This section describes the five kinds of groups in general.

Event group

The event group defines event indexes and controls the generation and notifications of the events triggered by the alarms defined in the alarm group and the private alarm group. The events can be handled in one of the following ways:

- Logging events in the event log table
- Sending traps to NMSs
- Both logging and sending traps
- No action

Alarm group

The RMON alarm group monitors specified alarm variables, such as statistics on a port. If the sampled value of the monitored variable is bigger than or equal to the upper threshold, an upper event is triggered; if the sampled value of the monitored variable is lower than or equal to the lower threshold, a lower event is triggered. The event is then handled as defined in the event group.

The following is how the system handles entries in the RMON alarm table:

- 1 Samples the alarm variables at the specified interval.
- 2 Compares the sampled values with the predefined threshold and triggers events if all triggering conditions are met.



If a sampled alarm variable overpasses the same threshold multiple times, only the first one can cause an alarm event. That is, the rising alarm and falling alarm are alternate.

Private alarm group

The private alarm group calculates the sampled values of alarm variables and compares the result with the defined threshold, thereby realizing a more comprehensive alarming function.

System handles the prialarm alarm table entry (as defined by the user) in the following ways:

- Periodically samples the prialarm alarm variables defined in the prialarm formula.
- Calculates the sampled values based on the prialarm formula.

- Compares the result with the defined threshold and generates an appropriate event.



If the count result overpasses the same threshold multiple times, only the first one can cause an alarm event. That is, the rising alarm and falling alarm are alternate.

History group

The history group controls the periodic statistical sampling of data, such as bandwidth utilization, number of errors, and total number of packets.

Note that each value provided by the group is a cumulative sum during a sampling period.

Ethernet statistics group

The statistics group monitors port utilization. It provides statistics about network collisions, CRC alignment errors, undersize/oversize packets, broadcasts, multicasts, bytes received, packets received, and so on.

After the creation of a valid event entry on a specified interface, the Ethernet statistics group counts the number of packets received on the current interface. The result of the statistics is a cumulative sum.

Configuring RMON

Configuration Prerequisites Before configuring RMON, configure the SNMP agent as described in “SNMP Configuration” on page 931.

Configuration Procedure Follow these steps to configure RMON:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create an event entry in the event table	rmon event <i>entry-number</i> [description <i>string</i>] { log log-trap <i>log-trapcommunity</i> none trap <i>trap-community</i> } [owner <i>text</i>]	Optional
Enter Ethernet interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Create an entry in the history table	rmon history <i>entry-number</i> buckets <i>number</i> interval <i>sampling-interval</i> [owner <i>text</i>]	Optional
Create an entry in the statistics table	rmon statistics <i>entry-number</i> [owner <i>text</i>]	Optional
Exit Ethernet interface view	quit	-
Create an entry in the alarm table	rmon alarm <i>entry-number</i> <i>alarm-variable</i> <i>sampling-interval</i> { absolute delta } rising-threshold <i>threshold-value1</i> <i>event-entry1</i> falling-threshold <i>threshold-value2</i> <i>event-entry2</i> [owner <i>text</i>]	Optional

To do...	Use the command...	Remarks
Create an entry in the private alarm table	rmon prialarm <i>entry-number</i> <i>prialarm-formula</i> <i>prialarm-des</i> <i>sampling-interval</i> { absolute changeratio delta } rising-threshold <i>threshold-value1</i> <i>event-entry1</i> falling-threshold <i>threshold-value2</i> <i>event-entry2</i> entrytype { forever cycle } [owner <i>text</i>]	Optional



- Two entries with the same configuration cannot be created. If the parameters of a newly created entry are identical to the corresponding parameters of an existing entry, the system considers their configurations the same and the creation fails. Refer to Table 75 for the parameters to be compared for different entries.
- The system limits the total number of all types of entries. When the total number of an entry reaches the maximum number of entries that can be created, the creation fails.
- When you create an entry in the history table, if the specified **buckets** number argument exceeds the history table size supported by the device, the entry will be created. However, the validated value of the **buckets** number argument corresponding with the entry is the history table size supported by the device.

Table 75 Restrictions on the configuration of RMON

Entry	Parameters to be compared
Event	Event description (description <i>string</i>), event type (log , trap , logtrap or none) and community name (<i>trap-community</i> or <i>log-trapcommunity</i>)
History	Sampling interval (interval <i>sampling-interval</i>)
Statistics	Only one statistics entry can be created on an interface.
Alarm	Alarm variable (<i>alarm-variable</i>), sampling interval (<i>sampling-interval</i>), sampling type (absolute or delta), rising threshold (<i>threshold-value1</i>) and falling threshold (<i>threshold-value2</i>)
Pri-alarm	Alarm variable formula (<i>alarm-variable</i>), sampling interval (<i>sampling-interval</i>), sampling type (absolute , changeratio or delta), rising threshold (<i>threshold-value1</i>) and falling threshold (<i>threshold-value2</i>)

Displaying and Maintaining RMON

To do...	Use the command...	Remarks
Display RMON statistics	display rmon statistics [<i>interface-type</i> <i>interface-number</i>]	Available in any view
Display RMON history information and the latest history sampling information	display rmon history [<i>interface-type</i> <i>interface-number</i>]	Available in any view
Display RMON alarm configuration information	display rmon alarm [<i>entry-number</i>]	Available in any view
Display RMON prialarm configuration information	display rmon prialarm [<i>entry-number</i>]	Available in any view
Display RMON events configuration information	display rmon event [<i>entry-number</i>]	Available in any view

To do...	Use the command...	Remarks
Display RMON event log information	display rmon eventlog [<i>event-number</i>]	Available in any view

RMON Configuration Example

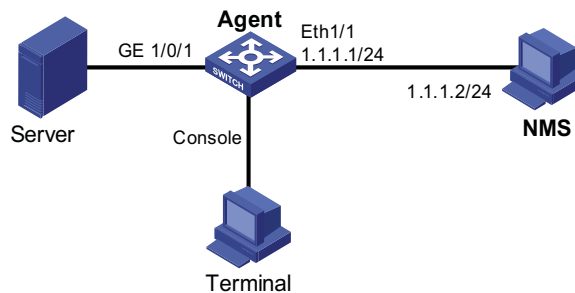
Network requirements

Agent is connected to a configuration terminal through its console port and to a remote NMS across the Internet.

Create an entry in the RMON Ethernet statistics table to gather statistics on GigabitEthernet 1/0/1, and logging is enabled after received bytes exceed the specified threshold.

Network diagram

Figure 276 Network diagram for RMON (on a switch)



Configuration procedure

Configure RMON to gather statistics for interface GigabitEthernet 1/0/1.

```
<Sysname> system-view
[Sysname] interface GigabitEthernet 1/0/1
[Sysname-GigabitEthernet 1/0/1] rmon statistics 1 owner user1-rmon
[Sysname-GigabitEthernet 1/0/1] quit
```

Display RMON statistics for interface GigabitEthernet 1/0/1.

```
<Sysname> display rmon statistics GigabitEthernet 1/0/1
Statistics entry 1 owned by user1-rmon is VALID.
Interface : GigabitEthernet1/0/1<ifIndex.1>
etherStatsOctets      : 0          , etherStatsPkts      : 0
etherStatsBroadcastPkts : 0          , etherStatsMulticastPkts : 0
etherStatsUndersizePkts : 0          , etherStatsOversizePkts : 0
etherStatsFragments   : 0          , etherStatsJabbers     : 0
etherStatsCRCAlignErrors : 0          , etherStatsCollisions  : 0
etherStatsDropEvents (insufficient resources): 0
Packets received according to length:
64      : 0          , 65-127 : 0          , 128-255 : 0
256-511: 0          , 512-1023: 0          , 1024-1518: 0
```

Create an event to start logging after the event is triggered.

```
<Sysname> system-view
[Sysname] rmon event 1 log owner 1-rmon
```

Configure an alarm group to sample received bytes on GigabitEthernet 1/0/1.
When the received bytes exceed the upper or below the lower limit, logging is
enabled.

```
[Sysname] rmon alarm 1 1.3.6.1.2.1.16.1.1.1.4.1 delta rising-threshold 1000 1 falling-threshold 100 1 owner 1-rmon
[Sysname] display rmon alarm 1
Alarm table 1 owned by 1-rmon is VALID.
Samples type          : delta
Variable formula      : 1.3.6.1.2.1.16.1.1.1.4.1<etherStatsOctets.1>
Sampling interval     : 10(sec)
Rising threshold      : 1000(linked with event 1)
Falling threshold     : 100(linked with event 1)
When startup enables  : risingOrFallingAlarm
Latest value          : 2552
```

NTP CONFIGURATION



The local clock of a Switch 4800G cannot be set as a reference clock. It can serve as a reference clock source to synchronize the clock of other devices only after it is synchronized.

When configuring NTP, go to these sections for information you are interested in:

- “NTP Overview” on page 947
- “NTP Configuration Task list” on page 953
- “Configuring the Operation Modes of NTP” on page 953
- “Configuring Optional Parameters of NTP” on page 956
- “Configuring Access-Control Rights” on page 957
- “Configuring NTP Authentication” on page 958
- “Displaying and Maintaining NTP” on page 960
- “NTP Configuration Examples” on page 960

NTP Overview

Defined in RFC 1305, the Network Time Protocol (NTP) synchronizes timekeeping among distributed time servers and clients. NTP runs over the User Datagram Protocol (UDP), using UDP port 123.

The purpose of using NTP is to keep consistent timekeeping among all clock-dependent devices within the network so that the devices can provide diverse applications based on the consistent time.

For a local system running NTP, its time can be synchronized by other reference sources and can be used as a reference source to synchronize other clocks.

Applications of NTP

An administrator can by no means keep synchronized time among all the devices within a network by changing the system clock on each station, because this is a huge amount of workload and cannot guarantee the clock precision. NTP, however, allows quick clock synchronization within the entire network while it ensures a high clock precision.

NTP is used when all devices within the network must be consistent in timekeeping, for example:

- In analysis of the log information and debugging information collected from different devices in network management, time must be used as reference basis.
- All devices must use the same reference clock in a charging system.

- To implement certain functions, such as scheduled restart of all devices within the network, all devices must be consistent in timekeeping.
- When multiple systems process a complex event in cooperation, these systems must use that same reference clock to ensure the correct execution sequence.
- For increment backup between a backup server and clients, timekeeping must be synchronized between the backup server and all the clients.

Advantages of NTP:

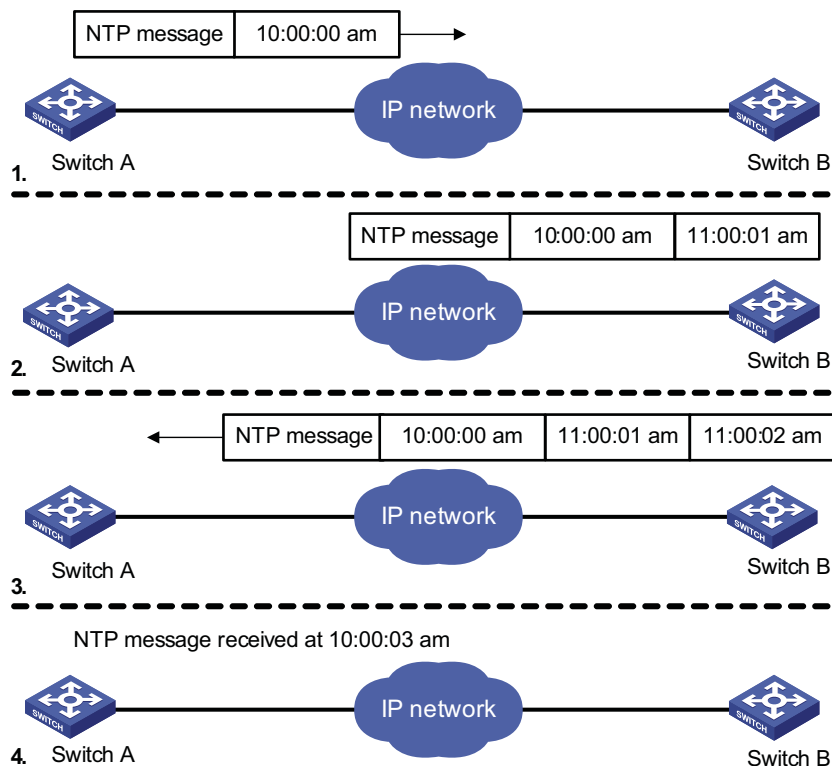
- NTP uses a stratum to describe the clock precision, and is able to synchronize time among all devices within the network.
- NTP supports access control and MD5 authentication.
- NTP can unicast, multicast or broadcast protocol messages.

How NTP Works

Figure 277 shows the basic work flow of NTP. Switch A and Switch B are interconnected over a network. They have their own independent system clocks, which need to be automatically synchronized through NTP. For an easy understanding, we assume that:

- Prior to system clock synchronization between Switch A and Switch B, the clock of Switch A is set to 10:00:00 am while that of Switch B is set to 11:00:00 am.
- Switch B is used as the NTP time server, namely Switch A synchronizes its clock to that of Switch B.
- It takes 1 second for an NTP message to travel from one switch to the other.

Figure 277 Basic work flow of NTP



The process of system clock synchronization is as follows:

- Switch A sends Switch B an NTP message, which is timestamped when it leaves Switch A. The time stamp is 10:00:00 am (T1).
- When this NTP message arrives at Switch B, it is timestamped by Switch B. The timestamp is 11:00:01 am (T2).
- When the NTP message leaves Switch B, Switch B timestamps it. The timestamp is 11:00:02 am (T3).
- When Switch A receives the NTP message, the local time of Switch A is 10:00:03 am (T4).

Up to now, Switch A has sufficient information to calculate the following two important parameters:

- The roundtrip delay of NTP message: $\text{Delay} = (T_4 - T_1) - (T_3 - T_2) = 2 \text{ seconds}$.
- Time difference between Switch A and Switch B: $\text{Offset} = ((T_2 - T_1) + (T_3 - T_4)) / 2 = 1 \text{ hour}$.

Based on these parameters, Switch A can synchronize its own clock to the clock of Switch B.

This is only a rough description of the work mechanism of NTP. For details, refer to RFC 1305.

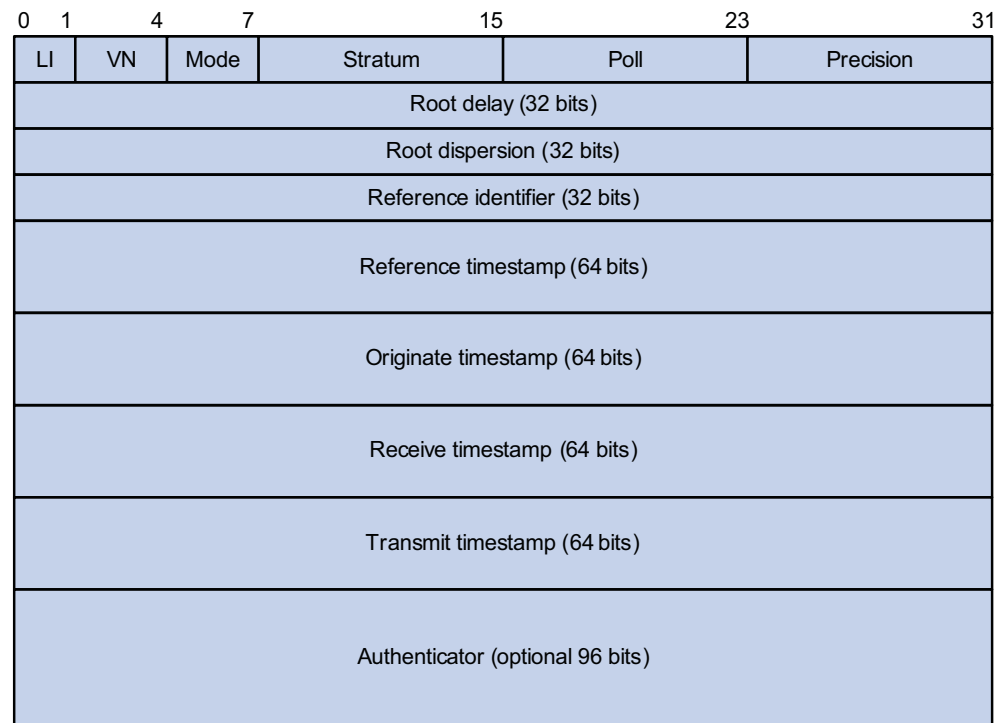
NTP Message Format

NTP uses two types of messages, clock synchronization message and NTP control message. An NTP control message is used in environments where network management is needed. As it is not a must for clock synchronization, it will not be discussed in this document.



All NTP messages mentioned in this document refer to NTP clock synchronization messages.

A clock synchronization message is encapsulated in a UDP message, in the format shown in Figure 278.

Figure 278 Clock synchronization message format

Main fields are described as follows:

- LI: 2-bit leap indicator. When set to 11, it warns of an alarm condition (clock unsynchronized); when set to any other value, it is not to be processed by NTP.
- VN: 3-bit version number, indicating the version of NTP. The latest version is version 3.
- Mode: a 3-bit code indicating the work mode of NTP. This field can be set to these values: 0 - reserved; 1 - symmetric active; 2 - symmetric passive; 3 - client; 4 - server; 5 - broadcast or multicast; 6 - NTP control message; 7 - reserved for private use.
- Stratum: an 8-bit integer indicating the stratum level of the local clock, with the value ranging from 1 to 16. The clock precision decreases from stratum 1 through stratum 16. A stratum 1 clock has the highest precision, and a stratum 16 clock is not synchronized and cannot be used as a reference clock.
- Poll: 8-bit signed integer indicating the poll interval, namely the maximum interval between successive messages.
- Precision: an 8-bit signed integer indicating the precision of the local clock.
- Root Delay: roundtrip delay to the primary reference source.
- Root Dispersion: the maximum error of the local clock relative to the primary reference source.
- Reference Identifier: Identifier of the particular reference source.
- Reference Timestamp: the local time at which the local clock was last set or corrected.
- Originate Timestamp: the local time at which the request departed the client for the service host.

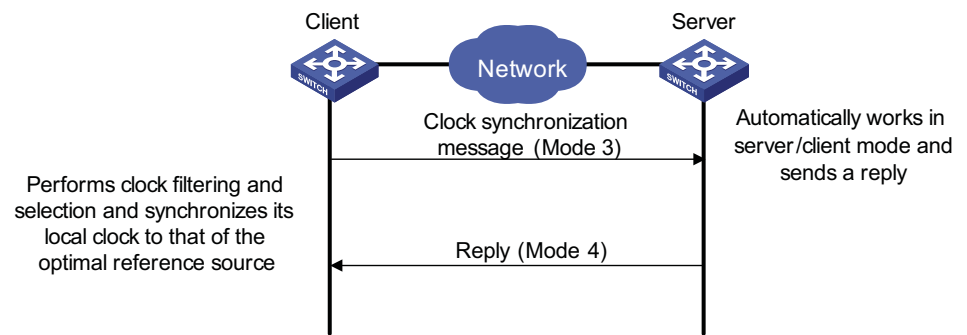
- Receive Timestamp: the local time at which the request arrived at the service host.
- Transmit Timestamp: the local time at which the reply departed the service host for the client.
- Authenticator: authentication information.

Operation Modes of NTP

Switches running NTP can implement clock synchronization in one of the following modes:

Server/client mode

Figure 279 Server/client mode

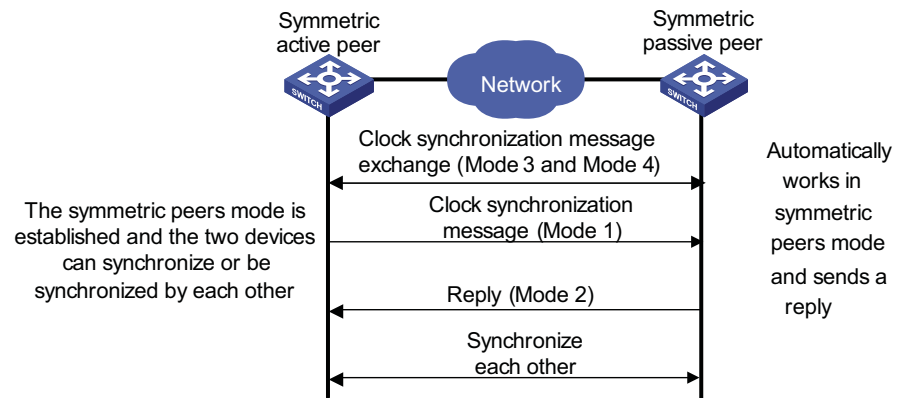


When working in the server/client mode, a client sends a clock synchronization message to servers, with the Mode field in the message set to 3 (client mode). Upon receiving the message, the servers automatically work in the server mode and send a reply, with the Mode field in the messages set to 4 (server mode). Upon receiving the replies from the servers, the client performs clock filtering and selection, and synchronizes its local clock to that of the optimal reference source.

In this mode, a client can be synchronized to a server, but not vice versa.

Symmetric peers mode

Figure 280 Symmetric peers mode

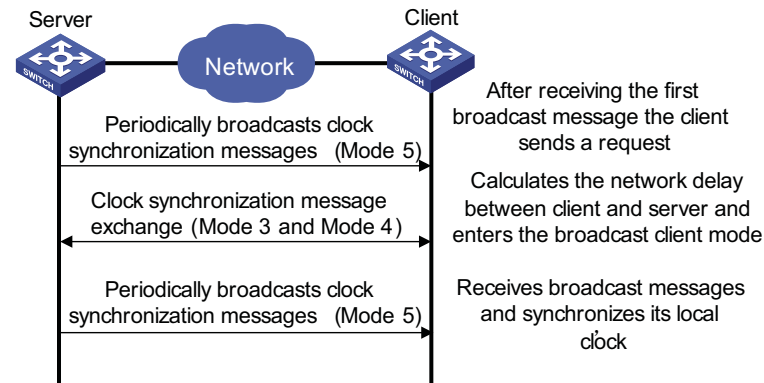


A switch working in the symmetric active mode periodically sends clock synchronization messages, with the Mode field in the message set to 1 (symmetric active); the switch that receives this message automatically enters the symmetric

passive mode and sends a reply, with the Mode field in the message set to 2 (symmetric passive). By exchanging messages, the symmetric peers mode is established between the two switches. Then, the two switches can synchronize, or be synchronized by, each other. If the clocks of both switches have been already synchronized, the switch whose local clock has a lower stratum level will synchronize the clock of the other switch.

Broadcast mode

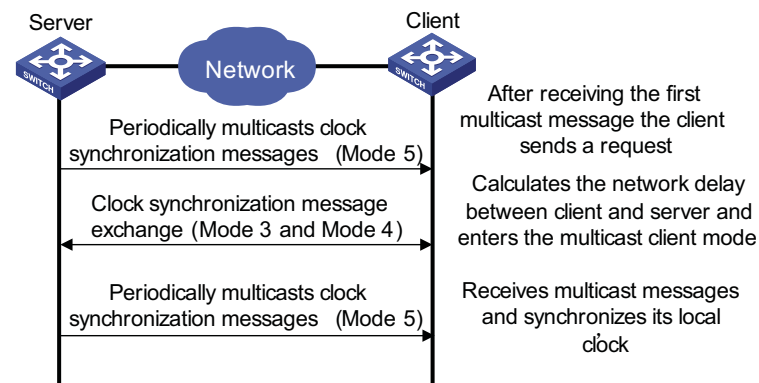
Figure 281 Broadcast mode



In the broadcast mode, a server periodically sends clock synchronization messages to the broadcast address 255.255.255.255, with the Mode field in the messages set to 5 (broadcast mode). Clients listen to the broadcast messages from servers. After a client receives the first broadcast message, the client and the server start to exchange messages, with the Mode field set to 3 (client mode) and 4 (server mode) to calculate the network delay between client and the server. Then, the client enters the broadcast client mode and continues listening to broadcast messages, and synchronizes its local clock based on the received broadcast messages.

Multicast mode

Figure 282 Multicast mode



In the multicast mode, a server periodically sends clock synchronization messages to the user-configured multicast address, or, if no multicast address is configured, to the default NTP multicast address 224.0.1.1, with the Mode field in the

messages set to 5 (multicast mode). Clients listen to the multicast messages from servers. After a client receives the first multicast message, the client and the server start to exchange messages, with the Mode field set to 3 (client mode) and 4 (server mode) to calculate the network delay between client and the server. Then, the client enters the multicast client mode and continues listening to multicast messages, and synchronizes its local clock based on the received multicast messages.



In symmetric peers mode, broadcast mode and multicast mode, the client (or the symmetric active peer) and the server (the symmetric passive peer) can work in the specified NTP working mode only after they exchange NTP messages with the Mode field being 3 (client mode) and the Mode field being 4 (server mode). During this message exchange process, NTP clock synchronization can be implemented.

NTP Configuration Task list

Complete the following tasks to configure NTP:

Task	Remarks
"Configuring the Operation Modes of NTP" on page 953	Required
"Configuring Optional Parameters of NTP" on page 956	Optional
"Configuring Access-Control Rights" on page 957	Optional
"Configuring NTP Authentication" on page 958	Optional

Configuring the Operation Modes of NTP

Switches can implement clock synchronization in one of the following modes:

- Server/client mode
- Symmetric mode
- Broadcast mode
- Multicast mode

For the server/client mode or symmetric mode, you need to configure only clients or symmetric-active peers; for the broadcast or multicast mode, you need to configure both servers and clients.



A single switch can have a maximum of 128 associations at the same time, including static associations and dynamic associations. A static association refers to an association that a user has manually created by using an NTP command, while a dynamic association is a temporary association created by the system during operation. A dynamic association will be removed if the system fails to receive messages from it over a specific long time. In the server/client mode, for example, when you carry out a command to synchronize the time to a server, the system will create a static association, and the server will just respond passively upon the receipt of a message, rather than creating an association (static or dynamic). In the symmetric mode, static associations will be created at the symmetric-active peer side, and dynamic associations will be created at the symmetric-passive peer side; In the broadcast or multicast mode, static associations will be created at the server side, and dynamic associations will be created at the client side.

Configuring NTP Server/Client Mode

For switches working in the server/client mode, you need to make configurations on the clients, and not on the servers.

Follow these steps to configure an NTP client:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Specify an NTP server for the switch	ntp-service unicast-server { <i>ip-address</i> <i>server-name</i> } [authentication-keyid <i>keyid</i> priority source-interface <i>interface-type interface-number</i> version <i>number</i>] *	Required No NTP server is specified by default.



- In the **ntp-service unicast-server** command, *ip-address* must be a host address, rather than a broadcast address, a multicast address or the IP address of the local clock.
- When the interface sending the NTP packet is specified by the **source-interface** argument, the source IP address of the NTP packet will be configured as the primary IP address of the specified interface.
- A switch can act as a server to synchronize the clock of other switches only after its clock has been synchronized. If the clock of a server has a stratum level higher than or equal to that of a client's clock, the client will not synchronize its clock to the server's.
- You can configure multiple servers by repeating the **ntp-service unicast-server** command. The clients will choose the optimal reference source.

Configuring the NTP Symmetric Mode

For switches working in the symmetric mode, you need to specify a symmetric-passive on a symmetric-active peer.

Following these steps to configure a symmetric-active switch:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Specify a symmetric-passive peer for the switch	ntp-service unicast-peer { <i>ip-address</i> <i>peer-name</i> } [authentication-keyid <i>keyid</i> priority source-interface <i>interface-type interface-number</i> version <i>number</i>] *	Required No symmetric-passive peer is specified by default.



- In the symmetric mode, you should use any NTP configuration command in "Configuring the Operation Modes of NTP" on page 953 to enable NTP; otherwise, a symmetric-passive peer will not process NTP packets from a symmetric-active peer.
- In the **ntp-service unicast-peer** command, *ip-address* must be a host address, rather than a broadcast address, a multicast address or the IP address of the local clock.
- When the interface used to send NTP messages is specified by the **source-interface** argument, the source IP address of the NTP message will be configured as the primary IP address of the specified interface.

- Typically, at least one of the symmetric-active and symmetric-passive peers has been synchronized; otherwise the clock synchronization will not proceed.
- You can configure multiple symmetric-passive peers by repeating the **ntp-service unicast-peer** command.

Configuring NTP Broadcast Mode

The broadcast server periodically sends NTP broadcast messages to the broadcast address 255.255.255.255. After receiving the messages, the switch working in NTP broadcast mode sends a reply and synchronizes its local clock.

For switches working in the broadcast mode, you need to configure both the server and clients. Because an interface need to be specified on the broadcast server for sending NTP broadcast messages and an interface also needs to be specified on each broadcast client for receiving broadcast messages, the NTP broadcast mode can be configured only in the specific interface view.

Configuring a broadcast client

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	Required Enter the interface used to receive NTP broadcast messages
Configure the switch to work in the NTP broadcast client mode	ntp-service broadcast-client	Required

Configuring the broadcast server

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	Enter the interface used to send NTP broadcast messages
Configure the switch to work in the NTP broadcast server mode	ntp-service broadcast-server [authentication-keyid <i>keyid</i> version <i>number</i>]*	Required



A broadcast server can synchronize broadcast clients only after its clock has been synchronized.

Configuring NTP Multicast Mode

The multicast server periodically sends NTP multicast messages to multicast clients, which send replies after receiving the messages and synchronize their local clocks.

For switches working in the multicast mode, you need to configure both the server and clients. The NTP multicast mode must be configured in the specific interface view.

Configuring a multicast client

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	Enter the interface used to receive NTP multicast messages
Configure the switch to work in the NTP multicast client mode	ntp-service multicast-client [<i>ip-address</i>]	Required

Configuring the multicast server

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	Enter the interface used to send NTP multicast message
Configure the switch to work in the NTP multicast server mode	ntp-service multicast-server [<i>ip-address</i>] [authentication-keyid <i>keyid</i> ttd <i>ttd-number</i> version <i>number</i>] *	Required



- A multicast server can synchronize broadcast clients only after its clock has been synchronized.
- You can configure up to 1024 multicast clients, among which 128 can take effect at the same time.

Configuring Optional Parameters of NTP

Configuring the Interface to Send NTP Messages

After you specify the interface used to send NTP messages, the source IP address of the NTP message will be configured as the primary IP address of the specified interface.

Following these steps to configure the interface used to send NTP messages:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the interface used to send NTP messages	ntp-service source-interface <i>interface-type interface-number</i>	Required



CAUTION: If you have specified an interface in the **ntp-service unicast-server** or **ntp-service unicast-peer** command, this interface will be used for sending NTP messages.

Disabling an Interface from Receiving NTP Messages

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-

Configuring the Maximum Number of Dynamic Sessions Allowed

To do...	Use the command...	Remarks
Disable the interface from receiving NTP messages	ntp-service in-interface disable	Required An interface is enabled to receive NTP messages by default

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the maximum number of dynamic sessions allowed to be established locally	ntp-service max-dynamic-sessions number	Required 100 by default

Configuring Access-Control Rights

With the following command, you can configure the NTP service access-control right to the local switch. There are four access-control rights, as follows:

- **query**: control query permitted. This level of right permits the peer switch to perform control query to the NTP service on the local switch but does not permit the peer switch to synchronize its clock to the local switch. The so-called “control query” refers to query of some states of the NTP service, including alarm information, authentication status, clock source information, and so on.
- **synchronization**: server access only. This level of right permits the peer switch to synchronize its clock to the local switch but does not permit the peer switch to perform control query.
- **server**: server access and query permitted. This level of right permits the peer switch to perform synchronization and control query to the local switch but does not permit the local switch to synchronize its clock to the peer switch.
- **peer**: full access. This level of right permits the peer switch to perform synchronization and control query to the local switch and also permits the local switch to synchronize its clock to the peer switch.

From the highest NTP service access-control right to the lowest one are **peer**, **server**, **synchronization**, and **query**. When a switch receives an NTP request, it will perform an access-control right match and will use the first matched right.

Configuration Prerequisites

Prior to configuring the NTP service access-control right to the local switch, you need to create and configure an ACL associated with the access-control right. For the configuration of ACL, refer to “ACL Overview” on page 835.

Configuration Procedure

Follow these steps to configure the NTP service access-control right to the local switch:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the NTP service access-control right to the local switch	ntp-service access { peer query server synchronization } acl-number	Required peer by default



The access-control right mechanism provides only a minimum degree of security protection for the system running NTP. A more secure method is identity authentication.

Configuring NTP Authentication

The NTP authentication feature should be enabled for a system running NTP in a network where there is a high security demand. This feature enhances the network security by means of client-server key authentication, which prohibits a client from synchronizing with a switch that has failed authentication.

Configuration Prerequisites

The configuration NTP authentication involves configuration tasks to be implemented on the client and on the server.

When configuring the NTP authentication feature, pay attention to the following principles:

- For all synchronization modes, when you enable the NTP authentication feature, you should configure an authentication key and specify it as a trusted key. Namely, the **ntp-service authentication enable** command must work together with the **ntp-service authentication-keyid** command and the **ntp-service reliable authentication-keyid** command. Otherwise, the NTP authentication function cannot be normally enabled.
- For the server/client mode or symmetric mode, you need to associate the specified authentication key on the client (symmetric-active peer if in the symmetric peer mode) with the corresponding NTP server (symmetric-passive peer if in the symmetric peer mode). Otherwise, the NTP authentication feature cannot be normally enabled.
- For the broadcast server mode or multicast server mode, you need to associate the specified authentication key on the broadcast server or multicast server with the corresponding NTP server. Otherwise, the NTP authentication feature cannot be normally enabled.
- For the server/client mode, if the NTP authentication feature has not been enabled for the client, the client can synchronize with the server regardless the NTP authentication feature has been enabled for the server or not.
- For all synchronization modes, the server side and the client side must be consistently configured.
- If the NTP authentication is enabled on a client, the client can be synchronized only to a server that can provide a trusted authentication key.

Configuration Procedure

Configuring NTP authentication for a client

Follow these steps to configure NTP authentication for a client:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable NTP authentication	ntp-service authentication enable	Required Disabled by default
Configure an NTP authentication key	ntp-service authentication-keyid <i>keyid</i> authentication-mode md5 <i>value</i>	Required No NTP authentication key by default

To do...	Use the command...	Remarks
Configure the key as a trusted key	ntp-service reliable authentication-keyid <i>keyid</i>	Required No authentication key is configured to be trusted by default
Associate the specified key with an NTP server	Server/client mode: ntp-service unicast-server { <i>ip-address</i> <i>server-name</i> } authentication-keyid <i>keyid</i> Symmetric peers mode: ntp-service unicast-peer { <i>ip-address</i> <i>peer-name</i> } authentication-keyid <i>keyid</i>	Required You can associate a non-existing key with an NTP server. To enable NTP authentication, you must configure the key and specify it as a trusted key after associating the key with the NTP server.



After you enable the NTP authentication feature for the client, make sure that you configure for the client an authentication key that is the same as on the server and specify that the authentication is trusted; otherwise, the client cannot be synchronized to the server.

Configuring NTP authentication for a server

Follow these steps to configure NTP authentication for a server:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable NTP authentication	ntp-service authentication enable	Required Disabled by default
Configure an NTP authentication key	ntp-service authentication-keyid <i>keyid</i> authentication-mode md5 <i>value</i>	Required No NTP authentication key by default
Configure the key as a trusted key	ntp-service reliable authentication-keyid <i>keyid</i>	Required No authentication key is configured to be trusted by default
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Associate the specified key with an NTP server	Broadcast server mode: ntp-service broadcast-server authentication-keyid <i>keyid</i> Multicast server mode: ntp-service multicast-server authentication-keyid <i>keyid</i>	Required You can associate a non-existing key with an NTP server. To enable NTP authentication, you must configure the key and specify it as a trusted key after associating the key with the NTP server.



The procedure of configuring NTP authentication on a server is the same as that on a client, and the same authentication key must be configured on both the server and client sides.

Displaying and Maintaining NTP

To do...	Use the command...	Remarks
View the information of NTP service status	display ntp-service status	Available in any view
View the information of NTP sessions	display ntp-service sessions [verbose]	Available in any view
View the brief information of the NTP servers from the local switch back to the primary reference source	display ntp-service trace	Available in any view

NTP Configuration Examples

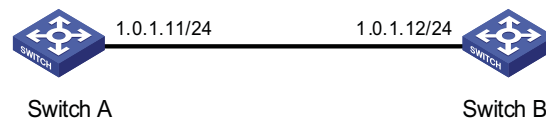
Configuring NTP Server/Client Mode

Network requirements

- The local clock of Switch A is to be used as a reference source, with the stratum level of 2.
- Switch B works in the server/client mode and Switch A is to be used as the NTP server of Switch B.

Network diagram

Figure 283 Network diagram for NTP server/client mode configuration



Configuration procedure

1 Configuration on Switch A:

Specify the local clock as the reference source, with the stratum level of 2.

```
<SwitchA> system-view
[SwitchA] ntp-service refclock-master 2
```

2 Configuration on Switch B:

View the NTP status of Switch B before clock synchronization.

```
<SwitchB> display ntp-service status
Clock status: unsynchronized
Clock stratum: 16
Reference clock ID: none
Nominal frequency: 100.0000 Hz
Actual frequency: 100.0000 Hz
Clock precision: 2^7
Clock offset: 0.0000 ms
Root delay: 0.00 ms
Root dispersion: 0.00 ms
Peer dispersion: 0.00 ms
Reference time: 00:00:00.000 UTC Jan 1 1900 (00000000.00000000)
```

Specify Switch A as the NTP server of Switch B so that Switch B is synchronized to Switch A.

```
<SwitchB> system-view
[SwitchB] ntp-service unicast-server 1.0.1.11
```

View the NTP status of Switch B after clock synchronization.

```
[SwitchB] display ntp-service status
Clock status: synchronized
Clock stratum: 3
Reference clock ID: 1.0.1.11
Nominal frequency: 100.0000 Hz
Actual frequency: 100.0000 Hz
Clock precision: 2^7
Clock offset: 0.0000 ms
Root delay: 31.00 ms
Root dispersion: 1.05 ms
Peer dispersion: 7.81 ms
Reference time: 14:53:27.371 UTC Apr 20 2007 (C6D94F67.5EF9DB22)
```

As shown above, Switch B has been synchronized to Switch A, and the clock stratum level of Switch B is 3, while that of Switch A is 2.

View the NTP session information of Switch B, which shows that an association has been set up between Switch B and Switch A.

```
[SwitchB] display ntp-service sessions
source      reference  stratum reach poll now offset delay disper
*****
[12345] 1.0.1.11 127.127.1.0 2 63 64 3 -75.5 31.0 16.5
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations : 1
```

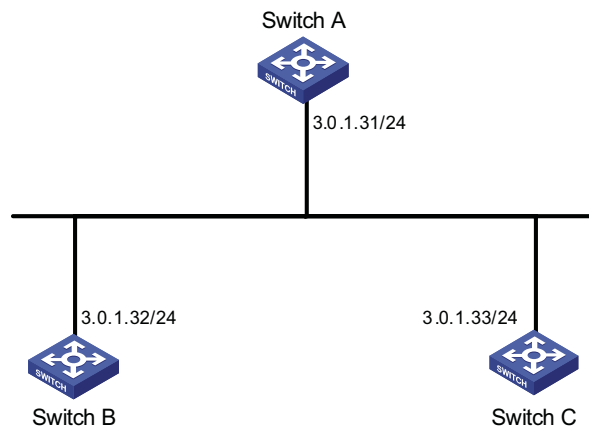
Configuring the NTP Symmetric Mode

Network requirements

- The local clock of Switch A is to be configured as a reference source, with the stratum level of 2.
- Switch B works in the client mode and Switch A is to be used as the NTP server of Switch B.
- Switch C works in the symmetric-active mode and Switch B will act as peer of Switch C. Switch C is the symmetric-active peer while Switch B is the symmetric-passive peer.

Network diagram

Figure 284 Network diagram for NTP symmetric peers mode configuration



Configuration procedure

1 Configuration on Switch A:

Specify the local clock as the reference source, with the stratum level of 2.

```
<SwitchA> system-view
[SwitchA] ntp-service refclock-master 2
```

2 Configuration on Switch B:

Specify Switch A as the NTP server of Switch B.

```
<SwitchB> system-view
[SwitchB] ntp-service unicast-server 3.0.1.31
```

3 Configuration on Switch C (after Switch B is synchronized to Switch A):

Specify the local clock as the reference source, with the stratum level of 1.

```
<SwitchC> system-view
[SwitchC] ntp-service refclock-master 1
```

Configure Switch B as a symmetric peer after local synchronization.

```
[SwitchC] ntp-service unicast-peer 3.0.1.32
```

In the step above, Switch B and Switch C are configured as symmetric peers, with Switch C in the symmetric-active mode and Switch B in the symmetric-passive mode. Because the stratum level of Switch C is 1 while that of Switch B is 3, Switch B is synchronized to Switch C.

View the NTP status of Switch B after clock synchronization.

```
[SwitchB] display ntp-service status
Clock status: synchronized
Clock stratum: 2
Reference clock ID: 3.0.1.33
Nominal frequency: 100.0000 Hz
Actual frequency: 100.0000 Hz
```

```

Clock precision: 2^7
Clock offset: -21.1982 ms
Root delay: 15.00 ms
Root dispersion: 775.15 ms
Peer dispersion: 34.29 ms
Reference time: 15:22:47.083 UTC Apr 20 2007 (C6D95647.153F7CED)

```

As shown above, Switch B has been synchronized to Switch C, and the clock stratum level of Switch B is 2, while that of Switch C is 1.

View the NTP session information of Switch B, which shows that an association has been set up between Switch B and Switch C.

```

[SwitchB] display ntp-service sessions
          source      reference  stra reach poll now  offset delay disper
*****
[245] 3.0.1.31 127.127.1.0    2   15   64  24  10535.0 19.6  14.5
[1234] 3.0.1.33 LOCL          1   14   64  27   -77.0  16.0  14.8
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations : 2

```

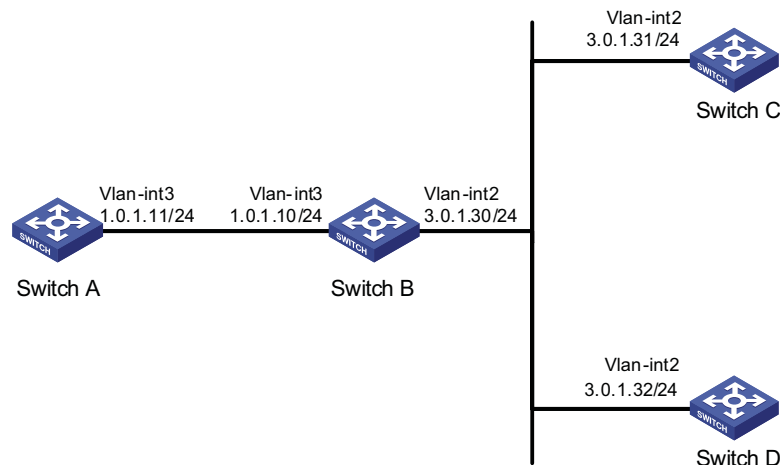
Configuring NTP Broadcast Mode

Network requirements

- Switch C's local clock is to be used as a reference source, with the stratum level of 2.
- Switch C works in the broadcast server mode and sends out broadcast messages from VLAN-interface 2.
- Switch D and Switch A work in the broadcast client mode and listen to broadcast messages through their respective VLAN-interface 2.

Network diagram

Figure 285 Network diagram for NTP broadcast mode configuration



Configuration procedure

1 Configuration on Switch C:

Specify the local clock as the reference source, with the stratum level of 2.

```

<SwitchC> system-view
[SwitchC] ntp-service refclock-master 2

```

Configure Switch C to work in the broadcast server mode and send broadcast messages through VLAN-interface 2.

```
[SwitchC] interface vlan-interface 2
[SwitchC-Vlan-interface2] ntp-service broadcast-server
```

1 Configuration on Switch D:

Configure Switch D to work in the broadcast client mode and receive broadcast messages on VLAN-interface 2.

```
<SwitchD> system-view
[SwitchD] interface vlan-interface 2
[SwitchD-Vlan-interface2] ntp-service broadcast-client
```

1 Configuration on Switch A:

Configure Switch A to work in the broadcast client mode and receive broadcast messages on VLAN-interface 3.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 3
[SwitchA-Vlan-interface3] ntp-service broadcast-client
```

Because Switch A and Switch C are on different subnets, Switch A cannot receive the broadcast messages from Switch C. Switch D gets synchronized upon receiving a broadcast message from Switch C.

View the NTP status of Switch D after clock synchronization.

```
[SwitchD] display ntp-service status
Clock status: synchronized
Clock stratum: 3
Reference clock ID: 3.0.1.31
Nominal frequency: 100.0000 Hz
Actual frequency: 100.0000 Hz
Clock precision: 2^7
Clock offset: 0.0000 ms
Root delay: 31.00 ms
Root dispersion: 8.31 ms
Peer dispersion: 34.30 ms
Reference time: 16:01:51.713 UTC Apr 20 2007 (C6D95F6F.B6872B02)
```

As shown above, Switch D has been synchronized to Switch A, and the clock stratum level of Switch D is 3, while that of Switch C is 2.

View the NTP session information of Switch D, which shows that an association has been set up between Switch D and Switch C.

```
[SwitchD] display ntp-service sessions
      source      reference      stra reach poll now      offset delay disper
*****
[1234] 3.0.1.31 127.127.1.0 2 254 64 62 -16.0 32.0 16.6
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations : 1
```

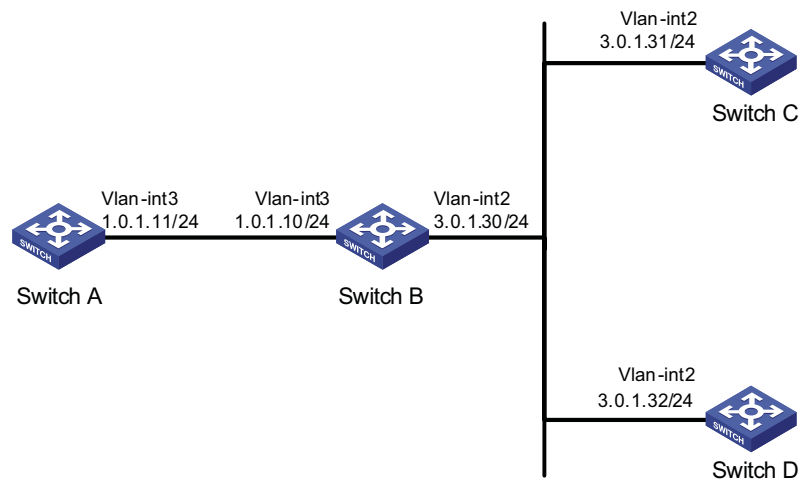

Configuring NTP Multicast Mode

Network requirements

- Switch C's local clock is to be used as a reference source, with the stratum level of 2.
- Switch C works in the multicast server mode and sends out multicast messages from VLAN-interface 2.
- Switch D and Switch A work in the multicast client mode and receive multicast messages through their respective VLAN-interface 2.

Network diagram

Figure 286 Network diagram for NTP multicast mode configuration



Configuration procedure

1 Configuration on Switch C:

Specify the local clock as the reference source, with the stratum level of 2.

```
<SwitchC> system-view
[SwitchC] ntp-service refclock-master 2
```

Configure Switch C to work in the multicast server mode and send multicast messages through VLAN-interface 2.

```
[SwitchC] interface vlan-interface 2
[SwitchC-Vlan-interface2] ntp-service multicast-server
```

2 Configuration on Switch D:

Configure Switch D to work in the multicast client mode and receive multicast messages on VLAN-interface 2.

```
<SwitchD> system-view
[SwitchD] interface vlan-interface 2
[SwitchD-Vlan-interface2] ntp-service multicast-client
```

Because Switch D and Switch C are on the same subnet, Switch D can receive the multicast messages from Switch C without being IGMP-enabled and can be synchronized to Switch C.

View the NTP status of Switch D after clock synchronization.

```
[SwitchD] display ntp-service status
Clock status: synchronized
Clock stratum: 3
Reference clock ID: 3.0.1.31
Nominal frequency: 100.0000 Hz
Actual frequency: 100.0000 Hz
Clock precision: 2^7
Clock offset: 0.0000 ms
Root delay: 31.00 ms
Root dispersion: 8.31 ms
Peer dispersion: 34.30 ms
Reference time: 16:01:51.713 UTC Apr 20 2007 (C6D95F6F.B6872B02)
```

As shown above, Switch D has been synchronized to Switch C, and the clock stratum level of Switch D is 3, while that of Switch C is 2.

View the NTP session information of Switch D, which shows that an association has been set up between Switch D and Switch C.

```
[SwitchD] display ntp-service sessions
      source      reference      stra reach poll now      offset delay disper
*****
[1234] 3.0.1.31 127.127.1.0 2 254 64 62 -16.0 31.0 16.6
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations : 1
```

3 Configuration on Switch B:

Because Switch A and Switch C are on different subnets, you must enable IGMP on Switch B before Switch A can receive multicast messages from Switch C.

Enable IP multicast routing and IGMP.

```
<SwitchB> system-view
[SwitchB] multicast routing-enable
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] pim dm
[SwitchB-Vlan-interface2] quit
[SwitchB] vlan 3
[SwitchB-vlan3] port GigabitEthernet 1/0/1
[SwitchB-vlan3] quit
[SwitchB] interface vlan-interface 3
[SwitchB-Vlan-interface3] igmp enable
[SwitchB-Vlan-interface3] quit
[SwitchB] interface GigabitEthernet 1/0/1
[SwitchB-GigabitEthernet1/0/1] igmp-snooping static-group 224.0.1.1 vlan 3
```

4 Configuration on Switch A:

Enable IP multicast routing and IGMP.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 3
```

Configure Switch A to work in the multicast client mode and receive multicast messages on VLAN-interface 3.

```
[SwitchA-Vlan-interface3] ntp-service multicast-client
```

View the NTP status of Switch A after clock synchronization.

```
[SwitchA] display ntp-service status
Clock status: synchronized
Clock stratum: 3
Reference clock ID: 3.0.1.31
Nominal frequency: 100.0000 Hz
Actual frequency: 100.0000 Hz
Clock precision: 2^7
Clock offset: 0.0000 ms
Root delay: 40.00 ms
Root dispersion: 10.83 ms
Peer dispersion: 34.30 ms
Reference time: 16:02:49.713 UTC Apr 20 2007 (C6D95F6F.B6872B02)
```

As shown above, Switch A has been synchronized to Switch C, and the clock stratum level of Switch A is 3, while that of Switch C is 2.

View the NTP session information of Switch A, which shows that an association has been set up between Switch A and Switch C.

```
[SwitchA] display ntp-service sessions
      source      reference      stra reach poll now      offset delay disper
*****
[1234] 3.0.1.31 127.127.1.0    2   255    64   26   -16.0  40.0  16.6
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations : 1
```



Refer to “Multicast Routing and Forwarding Configuration” on page 701 for detailed description of the multicast function.

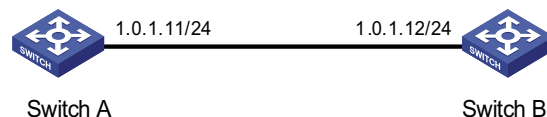
Configuring NTP Server/Client Mode with Authentication

Network requirements

- The local clock of Switch A is to be configured as a reference source, with the stratum level of 2.
- Switch B works in the client mode and Switch A is to be used as the NTP server of Switch B, with Switch B as the client.
- NTP authentication is to be enabled for Switch A and Switch B at the same time.

Network diagram

Figure 287 Network diagram for configuration of NTP server/client mode with authentication



Configuration procedure

1 Configuration on Switch A:

Specify the local clock as the reference source, with the stratum level of 2.

```
<SwitchA> system-view
[SwitchA] ntp-service refclock-master 2
```

2 Configuration on Switch B:

```
<SwitchB> system-view
```

```
# Enable NTP authentication on Switch B.
```

```
[SwitchB] ntp-service authentication enable
```

```
# Set an authentication key.
```

```
[SwitchB] ntp-service authentication-keyid 42 authentication-mode md5 aNiceKey
```

```
# Specify the key as key as a trusted key.
```

```
[SwitchB] ntp-service reliable authentication-keyid 42
```

```
# Specify Switch A as the NTP server.
```

```
[SwitchB] ntp-service unicast-server 1.0.1.11 authentication-keyid 42
```

Before Switch B can synchronize its clock to that of Switch A, you need to enable NTP authentication for Switch A.

Perform the following configuration on Switch A:

```
# Enable NTP authentication.
```

```
[SwitchA] ntp-service authentication enable
```

```
# Set an authentication key.
```

```
[SwitchA] ntp-service authentication-keyid 42 authentication-mode md5 aNiceKey
```

```
# Specify the key as key as a trusted key.
```

```
[SwitchA] ntp-service reliable authentication-keyid 42
```

```
# View the NTP status of Switch B after clock synchronization.
```

```
[SwitchB] display ntp-service status
Clock status: synchronized
Clock stratum: 3
Reference clock ID: 1.0.1.11
Nominal frequency: 100.0000 Hz
Actual frequency: 100.0000 Hz
Clock precision: 2^7
Clock offset: 0.0000 ms
Root delay: 31.00 ms
Root dispersion: 1.05 ms
Peer dispersion: 7.81 ms
Reference time: 14:53:27.371 UTC Apr 20 2007 (C6D94F67.5EF9DB22)
```

As shown above, Switch B has been synchronized to Switch A, and the clock stratum level of Switch B is 3, while that of Switch A is 2.

```
# View the NTP session information of Switch B, which shows that an association has been set up Switch B and Switch A.
```

```
[SwitchB] display ntp-service sessions
source      reference  strata reach poll now offset delay disper
*****
[12345] 1.0.1.11 127.127.1.0 2 63 64 3 -75.5 31.0 16.5
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations : 1
```

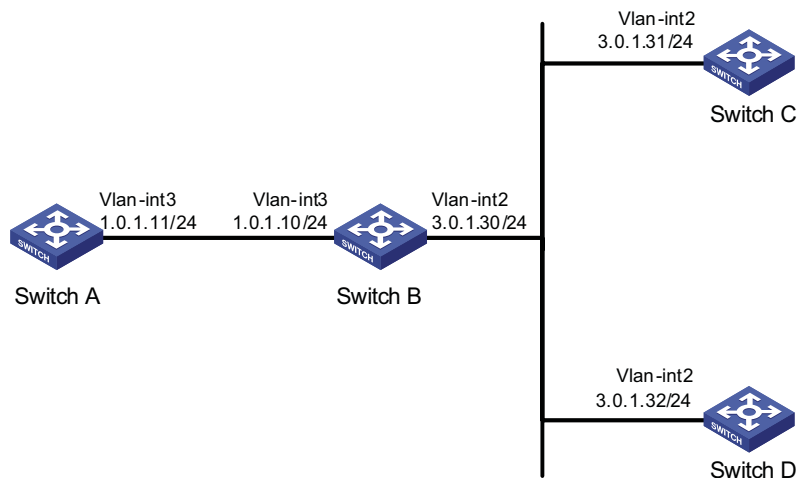
Configuring NTP Broadcast Mode with Authentication

Network requirements

- Switch C's local clock is to be used as a reference source, with the stratum level of 3.
- Switch C works in the broadcast server mode and sends out broadcast messages from VLAN-interface 2.
- Switch D works in the broadcast client mode and receives broadcast messages through VLAN-interface 2.
- NTP authentication is enabled on both Switch C and Switch D.

Network diagram

Figure 288 Network diagram for configuration of NTP broadcast mode with authentication



Configuration procedure

1 Configuration on Switch C:

Specify the local clock as the reference source, with the stratum level of 3.

```
<SwitchC> system-view
[SwitchC] ntp-service refclock-master 3
```

Configure NTP authentication

```
[SwitchC] ntp-service authentication enable
[SwitchC] ntp-service authentication-keyid 88 authentication-mode md5 123456
[SwitchC] ntp-service reliable authentication-keyid 88
```

Specify Switch C as an NTP broadcast server, and specify an authentication key.

```
[SwitchC] interface vlan-interface 2
[SwitchC-Vlan-interface2] ntp-service broadcast-server authentication-keyid 88
```

2 Configuration on Switch D:

Configure NTP authentication

```
<SwitchD> system-view
[SwitchD] ntp-service authentication enable
[SwitchD] ntp-service authentication-keyid 88 authentication-mode md5 123456
[SwitchD] ntp-service reliable authentication-keyid 88
```

Configure Switch D to work in the NTP broadcast client mode

```
[SwitchD] interface vlan-interface 2
[SwitchD-Vlan-interface2] ntp-service broadcast-client
```

Now, Switch D can receive broadcast messages through VLAN-interface 2, and Switch C can send broadcast messages through VLAN-interface 2. Upon receiving a broadcast message from Switch C, Switch D synchronizes its clock to that of Switch C.

View the NTP status of Switch D after clock synchronization.

```
[SwitchD] display ntp-service status
Clock status: synchronized
Clock stratum: 4
Reference clock ID: 3.0.1.31
Nominal frequency: 100.0000 Hz
Actual frequency: 100.0000 Hz
Clock precision: 2^7
Clock offset: 0.0000 ms
Root delay: 31.00 ms
Root dispersion: 8.31 ms
Peer dispersion: 34.30 ms
Reference time: 16:01:51.713 UTC Apr 20 2007 (C6D95F6F.B6872B02)
```

As shown above, Switch D has been synchronized to Switch B, and the clock stratum level of Switch D is 3, while that of Switch C is 2.

View the NTP session information of Switch D, which shows that an association has been set up between Switch D and Switch C.

```
[SwitchD] display ntp-service sessions
      source      reference      stra reach poll now      offset delay disper
*****
[1234] 3.0.1.31 127.127.1.0 3 254 64 62 -16.0 32.0 16.6
note: 1 source(master),2 source(peer),3 selected,4 candidate,5 configured
Total associations : 1
```

When configuring DNS, go to these sections for information you are interested in:

- “DNS Overview” on page 971
- “Configuring the DNS Client” on page 973
- “Configuring the DNS Proxy” on page 974
- “Displaying and Maintaining DNS” on page 974
- “DNS Configuration Examples” on page 975
- “Troubleshooting DNS Configuration” on page 980



This document only covers IPv4 DNS configurations. For introduction to IPv6 DNS configurations, refer to “Configuring IPv6 DNS” on page 515.

DNS Overview

Domain Name System (DNS) is a distributed database used by TCP/IP applications to translate domain names into corresponding IP addresses. With DNS, you can use easy-to-remember domain names in some applications and let the DNS server translate them into correct IP addresses.

There are two types of DNS services, static and dynamic. After a user specifies a name, the device checks the local static name resolution table for an IP address. If no IP address is available, it contacts the DNS server for dynamic name resolution, which takes more time than static name resolution. Therefore, some frequently queried name-to-IP address mappings are stored in the local static name resolution table to improve efficiency.

Static Domain Name Resolution

The static domain name resolution means setting up mappings between domain names and IP addresses. IP addresses of the corresponding domain names can be found in the static domain resolution table when you use applications such as telnet.

Dynamic Domain Name Resolution

Resolving procedure

Dynamic domain name resolution is implemented by querying the DNS server. The resolution procedure is as follows:

- 1 A user program sends a name query to the resolver of the DNS client.
- 2 The DNS resolver looks up the local domain name cache for a match. If a match is found, it sends the corresponding IP address back. If not, it sends a query to the DNS server.
- 3 The DNS server looks up the corresponding IP address of the domain name in its DNS database. If no match is found, it sends a query to a higher level DNS server. This process continues until a result, whether successful or not, is returned.

- 4 The DNS client returns the resolution result to the application after receiving a response from the DNS server.

Figure 289 Dynamic domain name resolution

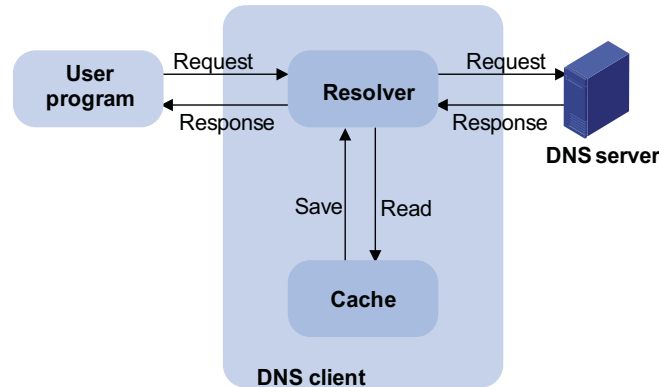


Figure 289 shows the relationship between the user program, DNS client, and DNS server.

The resolver and cache comprise the DNS client. The user program and DNS client can run on the same device or different devices, while the DNS server and the DNS client usually run on different devices.

Dynamic domain name resolution allows the DNS client to store latest mappings between domain names and IP addresses in the dynamic domain name cache. There is no need to send a request to the DNS server for a repeated query next time. The aged mappings are removed from the cache after some time, and latest entries are required from the DNS server. The DNS server decides how long a mapping is valid, and the DNS client gets the aging information from DNS messages.

DNS suffixes

The DNS client normally holds a list of suffixes which can be defined by users. It is used when the name to be resolved is incomplete. The resolver can supply the missing part. For example, a user can configure com as the suffix for aabbcc.com. The user only needs to type aabbcc to get the IP address of aabbcc.com. The resolver can add the suffix and delimiter before passing the name to the DNS server.

- If there is no dot in the domain name (for example, aabbcc), the resolver will consider this a host name and add a DNS suffix before query. If no match is found after all the configured suffixes are used respectively, the original domain name (for example, aabbcc) is used for query.
- If there is a dot in the domain name (for example, www.aabbcc), the resolver will directly use this domain name for query. If the query fails, the resolver adds a DNS suffix for another query.
- If the dot is at the end of the domain name (for example, aabbcc.com.), the resolver will consider it a fully qualified domain name (FQDN) and return the query result, successful or failed. Hence, the dot "." at the end of the domain name is called the terminating symbol.

Currently, the device supports static and dynamic DNS services.



If an alias is configured for a domain name on the DNS server, the device can resolve the alias into the IP address of the host.

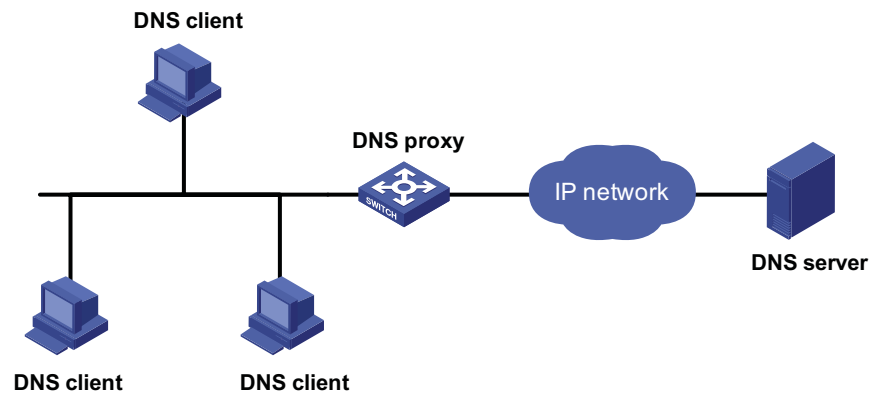
DNS Proxy Introduction to DNS proxy

A DNS proxy forwards DNS requests and replies between DNS clients and a DNS server.

As shown in Figure 290, a DNS client sends a DNS request to the DNS proxy, which forwards the request to the designated DNS server, and conveys the reply from the DNS server to the client.

The DNS proxy simplifies network management. When the DNS server address is changed, you only need to change the configuration on the DNS proxy instead of on each DNS client.

Figure 290 DNS proxy networking application



Operation of a DNS proxy

- 1 A DNS client considers the DNS proxy as the DNS server, and sends a DNS request to the DNS proxy, that is, the destination address of the request is the IP address of the DNS proxy.
- 2 The DNS proxy searches the local static domain name resolution table after receiving the request. If the requested information exists in the table, the DNS proxy returns a DNS reply to the client.
- 3 If the requested information does not exist in the static domain name resolution table, the DNS proxy sends the request to the designated DNS server for domain name resolution.
- 4 After receiving a reply from the DNS server, the DNS proxy forwards the reply to the DNS client.

Configuring the DNS Client

Configuring Static Domain Name Resolution

Follow these steps to configure static domain name resolution:

To do...	Use the command...	Remarks
Enter system view	system-view	--

To do...	Use the command...	Remarks
Configure a mapping between a host name and IP address in the static name resolution table	ip host <i>hostname ip-address</i>	Required Not configured by default.



The IP address you last assign to the host name will overwrite the previous one if there is any.

You may create up to 50 static mappings between domain names and IP addresses.

Configuring Dynamic Domain Name Resolution

Follow these steps to configure dynamic domain name resolution:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable dynamic domain name resolution	dns resolve	Required Disabled by default.
Specify a DNS server	dns server <i>ip-address</i>	Required Not specified by default
Configure a domain name suffix	dns domain <i>domain-name</i>	Optional Not configured by default



You may configure up to six DNS servers and ten DNS suffixes.

Configuring the DNS Proxy

Follow these steps to configure the DNS proxy:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable DNS proxy	dns proxy enable	Required Disabled by default.

Displaying and Maintaining DNS

To do...	Use the command...	Remarks
Display the static domain name resolution table	display ip host	Available in any view
Display DNS server information	display dns server [dynamic]	
Display domain name suffixes	display dns domain [dynamic]	Available in any view
Display the information of the dynamic domain name cache	display dns dynamic-host	
Display the DNS proxy table	display dns proxy table	
Clear the information of the dynamic domain name cache	reset dns dynamic-host	Available in user view

DNS Configuration Examples

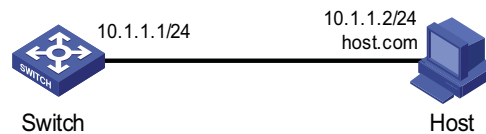
Static Domain Name Resolution Configuration Example

Network requirements

Switch uses the static domain name resolution to access Host with IP address 10.1.1.2 through domain name host.com.

Network diagram

Figure 291 Network diagram for static domain name resolution



Configuration procedure

Configure a mapping between host name host.com and IP address 10.1.1.2.

```
<Sysname> system-view
[Sysname] ip host host.com 10.1.1.2
```

Execute the **ping host.com** command to verify that the Switch can use the static domain name resolution to get the IP address 10.1.1.2 corresponding to host.com.

```
[Sysname] ping host.com
PING host.com (10.1.1.2):
 56 data bytes, press CTRL_C to break
  Reply from 10.1.1.2: bytes=56 Sequence=1 ttl=128 time=2 ms
  Reply from 10.1.1.2: bytes=56 Sequence=2 ttl=128 time=2 ms
  Reply from 10.1.1.2: bytes=56 Sequence=3 ttl=128 time=2 ms
  Reply from 10.1.1.2: bytes=56 Sequence=4 ttl=128 time=2 ms
  Reply from 10.1.1.2: bytes=56 Sequence=5 ttl=128 time=2 ms

--- host.com ping statistics ---
 5 packet(s) transmitted
 5 packet(s) received
 0.00% packet loss
 round-trip min/avg/max = 2/2/2 ms
```

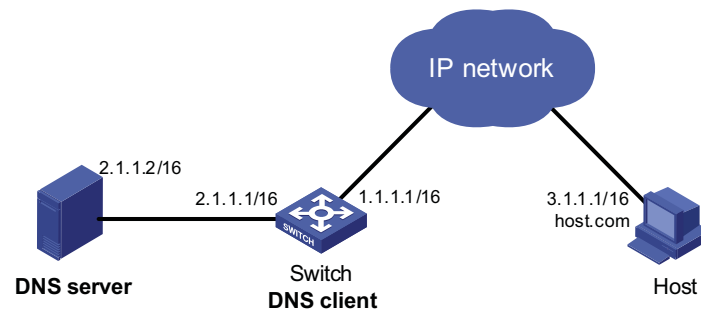
Dynamic Domain Name Resolution Configuration Example

Network requirements

- The IP address of the DNS server is 2.1.1.2/16 and the name suffix is com.
- Switch serving as a DNS client uses the dynamic domain name resolution and the suffix to access the host with the domain name host.com and the IP address 3.1.1.1/16.

Network diagram

Figure 292 Network diagram for dynamic domain name resolution



Configuration procedure



- Before performing the following configuration, make sure that there is a route between the device and the host, and configurations are done on both the device and the host. For the IP addresses of the interfaces, see Figure 292.
- This configuration may vary with different DNS servers. The following configuration is performed on a Windows 2000 server.
- Configure the DNS server

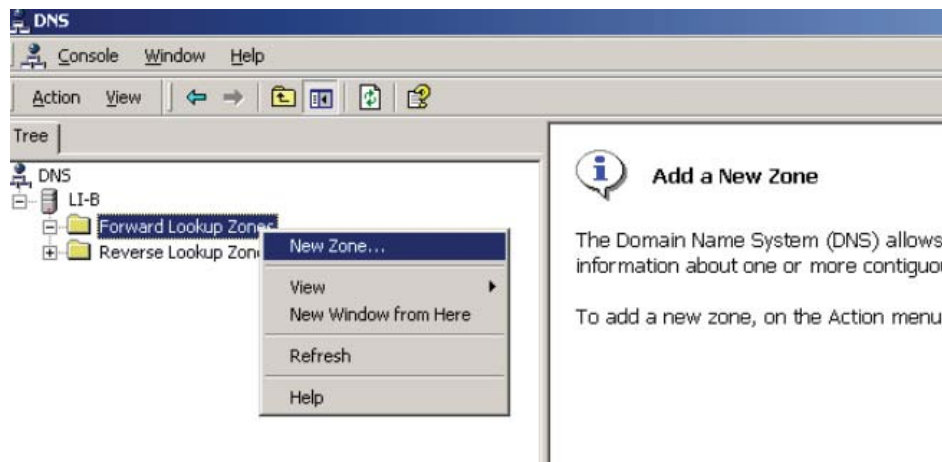
Enter DNS server configuration page.

Select **Start > Programs > Administrative Tools > DNS**.

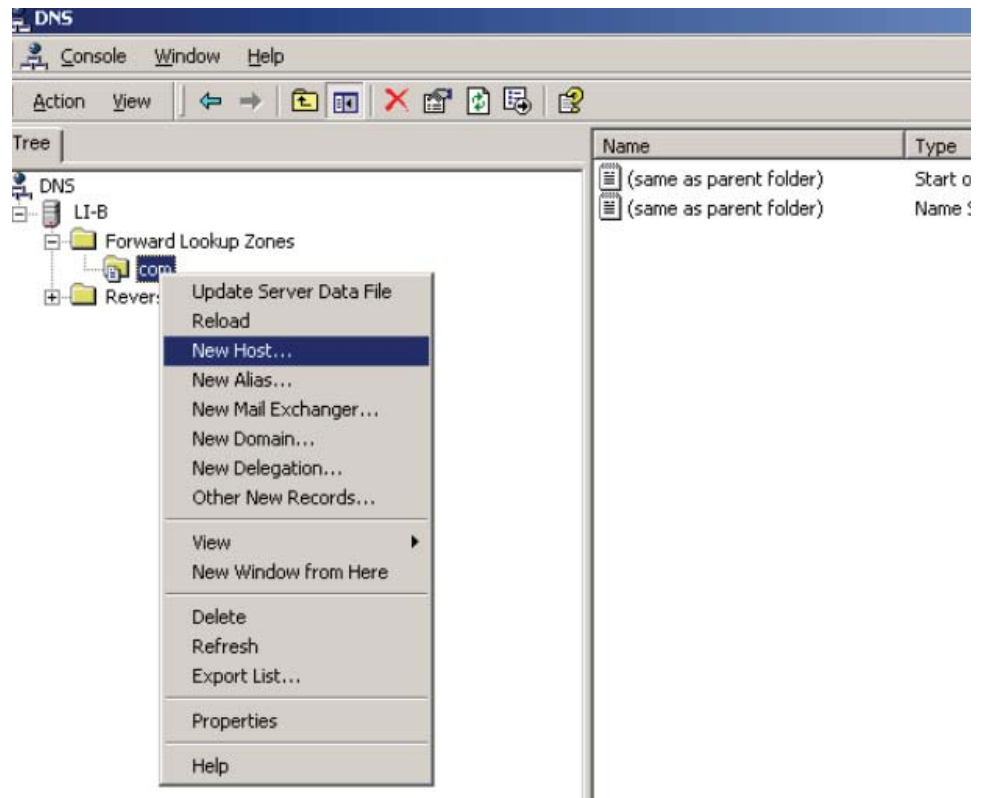
Create zone com.

In Figure 293, right click **Forward Lookup Zones**, select **New zone**, and then follow the instructions to create a new zone.

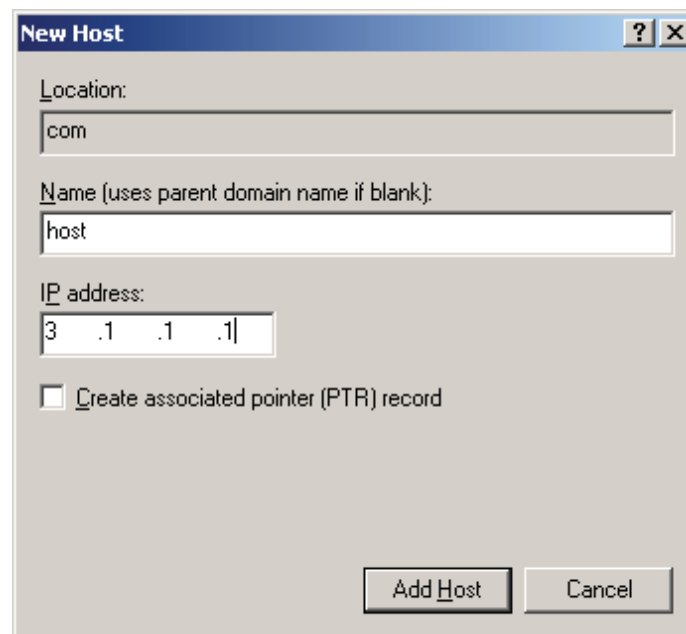
Figure 293 Create a zone



Create a mapping between the host name and IP address.

Figure 294 Add a host

In Figure 294, right click zone **com**, and then select **New Host** to bring up a dialog box as shown in Figure 295. Enter host name **host** and IP address **3.1.1.1**.

Figure 295 Add a mapping between domain name and IP address

1 Configure the DNS client

Enable dynamic domain name resolution.

```

<Sysname> system-view
[Sysname] dns resolve

# Specify the DNS server 2.1.1.2.

[Sysname] dns server 2.1.1.2

# Configure com as the name suffix.

[Sysname] dns domain com

```

2 Configuration verification

Execute the **ping host** command on the device to verify that the communication between the device and the host is normal and that the corresponding destination IP address is 3.1.1.1.

```

[Sysname] ping host
Trying DNS resolve, press CTRL_C to break
Trying DNS server (2.1.1.2)
PING host.com (3.1.1.1):
56 data bytes, press CTRL_C to break
  Reply from 3.1.1.1: bytes=56 Sequence=1 ttl=126 time=3 ms
  Reply from 3.1.1.1: bytes=56 Sequence=2 ttl=126 time=1 ms
  Reply from 3.1.1.1: bytes=56 Sequence=3 ttl=126 time=1 ms
  Reply from 3.1.1.1: bytes=56 Sequence=4 ttl=126 time=1 ms
  Reply from 3.1.1.1: bytes=56 Sequence=5 ttl=126 time=1 ms

--- host.com ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 1/1/3 ms

```

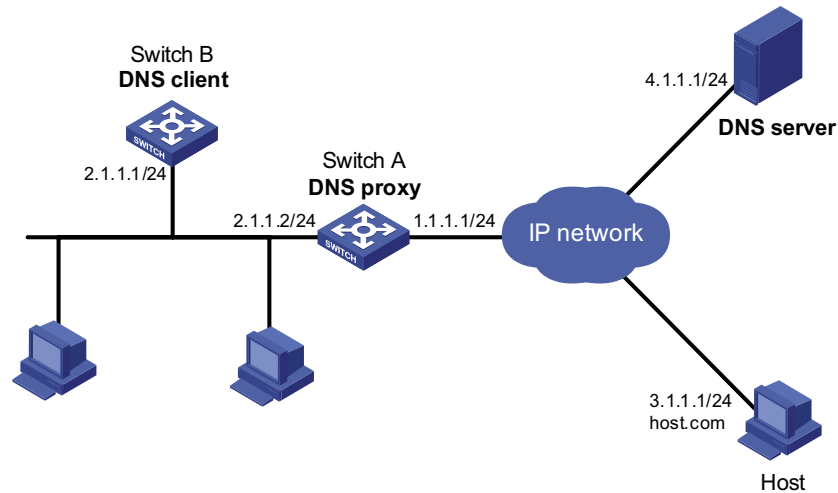
DNS Proxy Configuration Example

Network requirements

- Specify Switch A as the DNS server of Switch B (the DNS client).
- Switch A acts as a DNS proxy. The IP address of the real DNS server is 4.1.1.1.
- Switch B implements domain name resolution through Switch A.

Network diagram

Figure 296 Network diagram for DNS proxy



Configuration procedure



Before performing the following configuration, assume that Switch A, the DNS server, and the host are reachable to each other and the IP addresses of the interfaces are configured as shown in Figure 296.

1 Configure the DNS server

This configuration may vary with different DNS servers. When a Windows 2000 server acts as the DNS server, refer to “Dynamic Domain Name Resolution Configuration Example” on page 975 for related configuration information.

2 Configure the DNS proxy

Specify the DNS server 4.1.1.1.

```
<SwitchA> system-view
[SwitchA] dns server 4.1.1.1
```

Enable DNS proxy.

```
[SwitchA] dns proxy enable
```

3 Configure the DNS client

Enable the domain name resolution function.

```
<SwitchB> system-view
[SwitchB] dns resolve
```

Specify the DNS server 2.1.1.2.

```
[SwitchB] dns server 2.1.1.2
```

1 Configuration verification

Execute the **ping host.com** command on Switch B to verify that the host can be pinged after the host’s IP address 3.1.1.1 is resolved.

```
[SwitchB] ping host.com
Trying DNS resolve, press CTRL_C to break
Trying DNS server (2.1.1.2)
PING host.com (3.1.1.1):
56 data bytes, press CTRL_C to break
  Reply from 3.1.1.1: bytes=56 Sequence=1 ttl=126 time=3 ms
  Reply from 3.1.1.1: bytes=56 Sequence=2 ttl=126 time=1 ms
  Reply from 3.1.1.1: bytes=56 Sequence=3 ttl=126 time=1 ms
  Reply from 3.1.1.1: bytes=56 Sequence=4 ttl=126 time=1 ms
  Reply from 3.1.1.1: bytes=56 Sequence=5 ttl=126 time=1 ms

--- host.com ping statistics ---
  5 packet(s) transmitted
  5 packet(s) received
  0.00% packet loss
  round-trip min/avg/max = 1/1/3 ms
```

Troubleshooting DNS Configuration

Symptom

After enabling the dynamic domain name resolution, the user cannot get the correct IP address.

Solution

- Use the **display dns dynamic-host** command to verify that the specified domain name is in the cache.
- If there is no defined domain name, check that dynamic domain name resolution is enabled and the DNS client can communicate with the DNS server.
- If the specified domain name is in the cache, but the IP address is incorrect, check that the DNS client has the correct IP address of the DNS server.
- Verify the mapping between the domain name and IP address is correct on the DNS server.

79

FILE SYSTEM MANAGEMENT CONFIGURATION

When configuring the file system management, go to these sections for information you are interested in:

- "File System Management" on page 981
- "Configuration File Management" on page 985
- "Displaying and Maintaining Device Configuration" on page 989



Throughout this document, a filename can be entered as either of the following

- *A fully qualified filename with the path included to indicate a file under a specific path. The filename can be 1 to 135 characters in length.*
- *A short filename with the path excluded to indicate a file in the current path. The filename can be 1 to 91 characters in length.*

File System Management

This section covers these topics:

- "File System Overview" on page 981
- "Directory Operations" on page 981
- "File Operations" on page 982
- "Storage Device Operations" on page 983
- "File System Prompt Mode Setting" on page 983
- "File System Operations Example" on page 984

File System Overview

A major function of the file system is to manage storage devices. It allows you to perform operations such as directory create and delete, and file copy and display. If an operation, delete or overwrite for example, may cause problems such as data loss or corruption, the file system will ask you to confirm the operation by default.

Depending on the managed object, file system operations fall into "Directory Operations" on page 981, "File Operations" on page 982, "Storage Device Operations" on page 983, and "File System Prompt Mode Setting" on page 983.

Directory Operations

Directory operations include create, delete, display the current path, display specified directory or file information as shown in the following table:

To do...	Use the command...	Remarks
Create a directory	mkdir <i>directory</i>	Optional Available in user view
Remove a directory	rmdir <i>directory</i>	Optional Available in user view

To do...	Use the command...	Remarks
Display the current path	pwd	Optional Available in user view
Display files or directories	dir [/all] [file-url]	Optional Available in user view
Change the current path	cd <i>directory</i>	Optional Available in user view



- *The directory to be removed must be empty, meaning before you remove a directory, you must delete all the files and the subdirectory under this directory. For file deletion, refer to the **delete** command and for subdirectory deletion, refer to the **rmdir** command.*
- *After the execution of the **rmdir** command, the files in this directory will be automatically deleted for ever.*

File Operations

File operations include delete (removing files into the recycle bin), restore the deleted, permanently delete (deleting files from the recycle bin), display, rename, copy, and move files, and display specified directory or file information as shown in the following table:

To do...	Use the command...	Remarks
Remove a file to the recycle bin or delete it permanently	delete [/unreserved] <i>file-url</i>	Optional Available in user view
Restore a file from the recycle bin	undelete <i>file-url</i>	Optional Available in user view
Empty the recycle bin	reset recycle-bin [/force]	Optional Available in user view
Display the contents of a file	more <i>file-url</i>	Optional Currently only a .txt file can be displayed. Available in user view
Rename a file	rename <i>fileurl-source</i> <i>fileurl-dest</i>	Optional Available in user view
Copy a file	copy <i>fileurl-source</i> <i>fileurl-dest</i>	Optional Available in user view
Move a file	move <i>fileurl-source</i> <i>fileurl-dest</i>	Optional Available in user view
Display files or directories	dir [/all] [<i>file-url</i>]	Optional Available in user view
Enter system view	system-view	-
Execute the batch file	execute <i>filename</i>	Optional



*You can create a file by copying or downloading or using the **save** command.*

**CAUTION:**

- Empty the recycle bin timely with the **reset recycle-bin** command to save memory space.
- As the **delete /unreserved** file-url command deletes a file permanently and the action cannot be undone, use it with caution.
- The **execute** command cannot ensure the execution of each command. For example, if a certain command is not correctly configured, the system will omit this command and go to the next one. Therefore, each configuration command in a batch file must be a standard configuration command, meaning the valid configuration information which can be displayed with the **display current-configuration** command after this command is configured successfully; otherwise, this command may not be executed correctly.

Storage Device Operations**Naming rules**

Naming rules of the storage devices are as follows:

- If there is only one storage device of the same type on the device, the physical device name of the storage device is the storage device name.
- If there are multiple storage devices with the same type on the device, the physical device name of the storage device is composed of the storage device type and the serial number of the storage device. The serial number is displayed in English letters such as a, b or c.
- If storage device partitioning is supported on the device, the name of the partition device is composed of the physical device name and partition number. The serial numbers of partitions are displayed in numbers such as 0, 1 or 2.



Currently, the storage device on an Switch 4800G is the Flash only, which is named **flash**.

Memory space management

You can use the **fixdisk** command to restore the space of a storage device or the **format** command to format a specified storage device as shown in the following table:

To do...	Use the command...	Remarks
Restore the space of a storage device	fixdisk device	Optional Available in user view
Format a storage device	format device	Optional Available in user view

You may use the two commands when some space of a storage device becomes inaccessible due to abnormal operations for example.



CAUTION: When you format a storage device, all the files stored on it are erased and cannot be restored. In particular, if there is a startup configuration file on the storage device, formatting the storage device results in loss of the startup configuration file.

File System Prompt Mode Setting

The file system provides the following two prompt modes:

- **alert**: where the system warns you about operations that may bring undesirable consequence such as file corruption or data loss.
- **quiet**: where the system does not do that in any cases.

To prevent undesirable consequence resulted from misoperations, the **alert** mode is preferred.

To do...	Use the command...	Remarks
Enter system view	system-view	-
Set the operation prompt mode of the file system	file prompt { alert quiet }	Optional The default is alert .

File System Operations Example

Display the files and the subdirectory under the current directory.

```
<Sysname> dir
Directory of flash:/

 0  drw-          -  Feb 16 2006 11:45:36  logfile
 1  -rw-          1218 Feb 16 2006 11:46:19  config.cfg
 2  drw-          -  Feb 16 2006 15:20:27  test
 3  -rw-        184108 Feb 16 2006 15:30:20  aaa.bin

14605 KB total (6890 KB free)
```

Create a new folder called **mytest** under the test directory.

```
<Sysname> cd test
<Sysname> mkdir mytest
%Created dir flash:/test/mytest.
```

Display the current working directory.

```
<Sysname> pwd
flash:/test
```

Display the files and the subdirectory under the test directory.

```
<Sysname> dir
Directory of flash:/test/

 0  drw-          -  Feb 16 2006 15:28:14  mytest

2540 KB total (2519 KB free)
```

Return to the upper directory.

```
<Sysname> cd ..
```

Display the current working directory.

```
<Sysname> pwd
flash:/
```

Configuration File Management

The device provides the configuration file management function with a user-friendly operating interface for you to manage the configuration files conveniently.

This section covers these topics:

- "Configuration File Overview" on page 985
- "Saving the Current Configuration" on page 986
- "Erasing the Startup Configuration File" on page 987
- "Specifying a Configuration File for Next Startup" on page 988
- "Backing up/Restoring the Configuration File for Next Startup" on page 988

Configuration File Overview

A configuration file saves the device configurations in command lines in text format. You can view configuration information conveniently through the configuration files.

Types of configuration

The configuration of a device falls into two types:

- Saved configuration, a configuration file used for initialization. If this file does not exist, the default parameters are used.
- Current configuration, which refers to the user's configuration during the operation of a device. This configuration is stored in the flash. It is removed when the device is rebooting.

Format of configuration file

Configuration files are saved as text files. They:

- Save configuration in the form of commands.
- Save only non-default configuration settings.
- List commands in sections by view in this view order: system, interface, routing protocol, and so on. Sections are separated with one or multiple blank lines or comment lines that start with a pound sign (#).
- End with a return.

Main/backup attribute of the configuration file

A main configuration file and a backup configuration file can exist simultaneously if the device supports main/backup configuration file attribute. As such, when the main configuration file is missing or damaged, the backup file can be used instead. This increases the safety and reliability of the file system compared with the device that only supports one configuration file. You can configure a file to have both the main and backup attributes, but only one file of either main or backup attribute is allowed on a device.

The following three situations are concerned with the main/backup attribute:

- When saving the current configuration, you can specify the file to be a main or backup or normal configuration file.

- When removing a configuration file from a device, you can specify to remove the main or backup configuration file. Or, if it is a file having both the main and backup attributes, you can specify to erase the main or backup attribute of the file.
- When setting the configuration file for next startup, you can specify the main/backup attribute of the file.

Startup with the configuration file

The following steps are taken during system startup:

- 1 If the main configuration file exists, the device initializes with this configuration.
- 2 If the main configuration file does not exist but the backup configuration file exists, the device initializes with the backup configuration.
- 3 If neither the main nor the backup configuration file exists, the device will:
- 4 Initialize with the default configuration file if it exists;
- 5 Or initialize with empty configuration if the default configuration file does not exist.

Saving the Current Configuration

You can modify the configuration on your device at the command line interface (CLI). To use the modified configuration for your subsequent startups, you must save it (using the **save** command) as a configuration file.

Modes in saving the configuration

- Fast saving mode. This is the mode when you use the **save** command without the **safely** keyword. The mode saves the file quicker but is likely to lose the original configuration file if the device reboots or the power fails during the process.
- Safe mode. This is the mode when you use the **save** command with the **safely** keyword. The mode saves the file slower but can retain the configuration file in the device even if the device reboots or the power fails during the process.



CAUTION: Device reboot or the power failure during configuration file saving may result in loss of the configuration file for next startup. In this case, the device should be started with empty configuration and after the device starts, you need to re-specify a configuration file for next startup. Refer to “Specifying a Configuration File for Next Startup” on page 988 for details.

Attributes of the configuration file when main/backup attribute is supported

- Main attribute. When you use the **save [safely] [main]** command to save the current configuration, the configuration file you get has main attribute. If this configuration file already exists and has backup attribute, the file will have both main and backup attributes after execution of this command. If the filename you entered is different from that existing in the system, this command will erase its main attribute to allow only one main attribute configuration file in the device.
- Backup attribute. When you use the **save [safely] backup** command to save the current configuration, the configuration file you get has backup attribute. If this configuration file already exists and has main attribute, the file will have both main and backup attributes after execution of this command. If the

filename you entered is different from that existing in the system, this command will erase its backup attribute to allow only one backup attribute configuration file in the device.

- Normal attribute. When you use the **save** *file-name* command to save the current configuration, the configuration file you get has normal attribute if it is not an existing file. Otherwise, the attribute is the original attribute of the file.

Follow the step below to save the current configuration:

To do...	Use the command...	Remarks
Save the current configuration	save [<i>file-name</i> [safely] [backup main]]	Required Available in any view



- *Fast saving mode is suitable for environments where power supply is stable. The safe mode, however, is preferred where stable power supply is unavailable or remote maintenance is involved.*
- *The extension name of the configuration file must be .cfg.*
- *If you press <Enter> after entering the **save** command, you can save the configuration file in an interactive way. In this way, you can use the default path or enter a filename to specify a new path, but the suffix of the filename must be ".cfg".*
- *In interactive mode, if you use the non-default path (that is, entering a new filename), the system sets the file as the main configuration file for next startup*

Erasing the Startup Configuration File

With the configuration file erased, your device will boot up with the default configuration next time it is powered on.

You may need to erase the configuration file for one of these reasons:

- After you upgrade software, the original configuration file does not match the new software.
- The startup configuration file is corrupted or not the one you need.

When main/backup attributes are supported, the following two situations exist:

- While the **reset saved-configuration** [**main**] command erases the configuration file with main attribute, it only deletes the main attribute of a configuration file having both main and backup attribute.
- While the **reset saved-configuration backup** command erases the configuration file with backup attribute, it only deletes the backup attribute of a configuration file having both main and backup attribute.

Follow the step below to erase the configuration file:

To do...	Use the command...	Remarks
Erase the startup configuration file from the storage device	reset saved-configuration [backup main]	Required Available in user view



CAUTION: This command will permanently delete the configuration file from the device. Use it with caution.

Specifying a Configuration File for Next Startup

You can assign main or backup attribute to the configuration file for next startup when main/backup attributes are supported on your device.

Assigning main attribute to the configuration file for next startup

- If you save the current configuration to the main configuration file, the system will automatically set the file as the main startup configuration file.
- You can also use the **startup saved-configuration *cfgfile* main** command to set the file as main startup configuration file.

Assigning backup attribute to the configuration file for next startup

- If you save the current configuration to the backup configuration file, the system will automatically set the file as the backup startup configuration file.
- You can also use the **startup saved-configuration *cfgfile* backup** command to set the file as backup startup configuration file.

Follow the step below to specify a configuration file for next startup:

To do...	Use the command...	Remarks
Specify a configuration file for next startup	startup saved-configuration <i>cfgfile</i> [backup main]	Required Available in user view



CAUTION: The configuration file must use “.cfg” as its extension name and the startup configuration file must be saved under the root directory of the device.

Backing up/Restoring the Configuration File for Next Startup

Backup/restore function overview

The backup/restore function allows you to backup or restore a configuration file for next startup through operations at the CLI. TFTP is used for intercommunication between the device and the server. The backup function enables you to backup a configuration file to the TFTP server, while the restore function enables you to download the configuration file from the TFTP server for next startup.



For 3Com Switch 4800G Family Ethernet switches, the file to be backed up or restored is the main configuration file for next startup.

Backing up the configuration file for next startup

To do...	Use the command...	Remarks
Back up the configuration file for next startup	backup startup-configuration to <i>dest-addr</i> [<i>filename</i>]	Required Available in user view



Before backup, you should

- *Ensure that the server is reachable, the server is enabled with TFTP service, and the client has permission to read and write.*

- Use the **display startup** command (in user view) to verify if you have set the startup configuration file, and use the **dir** command to verify if this file exists. If the file is set as NULL or does not exist, the backup will be unsuccessful.

Restoring the startup configuration file

To do...	Use the command...	Remarks
Restore the startup configuration file	restore startup-configuration from src-addr filename	Required Available in user view



- Before restoring a configuration file, you should ensure that the server is reachable, the server is enabled with TFTP service, and the client has permission to read and write.
- After the command is successfully executed, you can use the **display startup** command (in user view) to verify if the filename of the startup configuration file is the same with the filename argument, and use the **dir** command to verify if the restored file exists.

Displaying and Maintaining Device Configuration

To do...	Use the command...	Remarks
Display the configuration file saved in the storage device	display saved-configuration [by-linenum]	Available in any view
Display the configuration file used for this and next startup	display startup	Available in any view
Display the validated configuration in current view	display this [by-linenum]	Available in any view
Display current configuration	display current-configuration [[configuration [configuration] controller interface [interface-type] [interface-number]] [by-linenum] [{ begin include exclude } text]]	Available in any view



For detailed description of the **display this** and **display current-configuration** commands, refer to “Displaying and Maintaining Device Management Configuration” on page 1043.

80

FTP CONFIGURATION

When configuring FTP, go to these sections for information you are interested in:

- “FTP Overview” on page 991
- “Configuring the FTP Client” on page 992
- “Configuring the FTP Server” on page 996
- “Displaying and Maintaining FTP” on page 999

FTP Overview

Introduction to FTP

The File Transfer Protocol (FTP) is an application layer protocol for sharing files between server and client over a TCP/IP network.

FTP uses TCP ports 20 and 21 for file transfer. Port 20 is used to transmit data, and port 21 to transmit control commands. Refer to RFC 959 for details of FTP basic operation.

FTP transmits files in two modes:

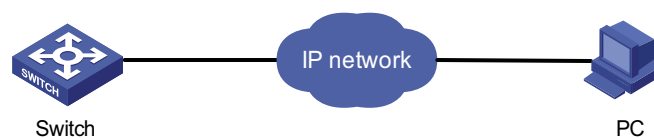
- Binary mode for program file transmission
- ASCII mode for text file transmission

Implementation of FTP

FTP adopts the server/client model. Your switch can function either as client or as server (as shown in Figure 297). They work in the following way:

- When the switch serves as the FTP client, a PC user first telnets or connects to the switch through an emulation program, then executes the **ftp** command to establish the connection to the remote FTP server, and gain access to the files on the server. If the remote FTP server supports anonymous FTP, the device can log onto it directly; if not, the device must obtain FTP username and password first to log onto the remote FTP server.
- When the switch serves as the FTP server, it must be configured with an IP address so that a user running FTP client program can access it. For the sake of security, the switch does not support anonymous FTP. Therefore, you must use an authenticated username and password. By default, authenticated users can access the root directory of the switch.

Figure 297 Network diagram for FTP



**CAUTION:**

- The FTP function is available when a route exists between the FTP server and the FTP client.
- When a device serving as the FTP server logs onto the device using IE, some IE functions are not supported because multiple user connections are established, and the device supports only one connection currently.

Configuring the FTP Client

Establishing an FTP Connection

To access an FTP server, the FTP client must connect with it. Two ways are available for the connection: using the **ftp** command to establish the connection directly; using the **open** command in FTP client view.

Multiple routes may exist for the FTP client to successfully access the FTP server. You can specify one by configuring the source address of the packets of the FTP client to meet the requirement of the security policy of the FTP client. You can configure the source address by configuring the source interface or source IP address. The primary IP address configured on the source interface is the source address of the transmitted packets. The source address of the transmitted packets is selected following these rules:

- If no source address of the FTP client is specified, a device uses the IP address of the interface determined by the routing protocol as the source IP address to communicate with an FTP server.
- If the source address is specified with the **ftp client source** or **ftp** command, this source address is used to communicate with an FTP server.
- If the source address is specified with the **ftp client source** command and then with the **ftp** command, the address specified with the latter one is used to communicate with an FTP server.

The source address specified with the **ftp client source** command is valid for all **ftp** connections and the source address specified with the **ftp** command is valid only for the current FTP connection.

Follow these steps to establish an FTP connection (In IPv4 networking):

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the source address of the FTP client	ftp client source { interface interface-type interface-number ip source-ip-address }	Optional A device uses the IP address of the interface determined by the routing protocol as the source IP address to communicate with the FTP server by default.
Exit to system view	quit	-

To do...	Use the command...	Remarks
Log onto the remote FTP server directly in user view	ftp [<i>server-address</i> [<i>service-port</i>] [source { interface <i>interface-type</i> <i>interface-number</i> ip <i>source-ip-address</i> }]]	Use either approach. Available in user view
Log onto the remote FTP server indirectly in FTP client view	ftp open <i>server-address</i> [<i>service-port</i>]	



- If no primary IP address is configured on the source interface, the FTP connection fails.
- If you use the **ftp client source** command to first configure the source interface and then the source IP address of the transmitted packets, the new source IP address will overwrite the current one, and vice versa.

Follow these steps to establish an FTP connection (In IPv6 networking):

To do...	Use the command...	Remarks
Log onto the remote FTP server directly in user view	ftp ipv6 [<i>server-address</i> [<i>service-port</i>] [source ipv6 <i>source-ipv6-address</i>] [-i <i>interface-type</i> <i>interface-number</i>]]	Use either approach. Available in user view
Log onto the remote FTP server indirectly in FTP client view	ftp ipv6 open ipv6 <i>server-address</i> [<i>service-port</i>] [-i <i>interface-type</i> <i>interface-number</i>]	

Configuring the FTP Client

After a device serving as the FTP client has established a connection with the FTP server (For establishing FTP connection, refer to “Establishing an FTP Connection” on page 992.), the device can perform the following operations for the authorized directory:

To do...	Use the command...	Remarks
Display help information of FTP-related commands supported by the remote FTP server	remotehelp [<i>protocol-command</i>]	Optional
Enable information display in a detailed manner	verbose	Optional Enabled by default
Use other username to relog after logging onto the FTP server successfully	user <i>username</i> [<i>password</i>]	Optional
Enable FTP client debugging	debugging	Optional Disabled by default
Set the file transfer mode to ASCII	ascii	Optional ASCII by default
Set the file transfer mode to binary	binary	Optional ASCII by default
Change the working path on the remote FTP server	cd <i>pathname</i>	Optional
Exit the current directory and enter the upper level directory	cdup	Optional

To do...	Use the command...	Remarks
Display files/directories information on the FTP server	dir [<i>remotefile</i> [<i>localfile</i>]]	Optional
Check files/directories on the FTP server	ls [<i>remotefile</i> [<i>localfile</i>]]	Optional
Download a file from the FTP server	get <i>remotefile</i> [<i>localfile</i>]	Optional
Upload a file to the FTP server	put <i>localfile</i> [<i>remotefile</i>]	Optional
View the working directory of the remote FTP server	pwd	Optional
Find the working path of the FTP client	lcd	Optional
Create a directory on the FTP server	mkdir <i>directory</i>	Optional
Set the data transfer mode to passive	passive	Optional Passive by default
Delete specified file on the FTP server	delete <i>remotefile</i>	Optional
Delete specified directory on the FTP server	rmdir <i>directory</i>	Optional
Disconnect with the FTP server without exiting the FTP client view	disconnect	Optional Equal to the close command
Disconnect with the FTP server without exiting the FTP client view	close	Optional Equal to the disconnect command
Disconnect with the FTP server and exit to user view	bye	Optional
Terminate the connection with the remote FTP server, and exit to user view	quit	Optional Available in FTP client view, equal to the bye command



- *FTP uses two modes for file transfer: ASCII mode and binary mode.*
- *The **ls** command can only display the file/directory name, while the **dir** command can display more information, such as the size and date of creation of files or directories.*

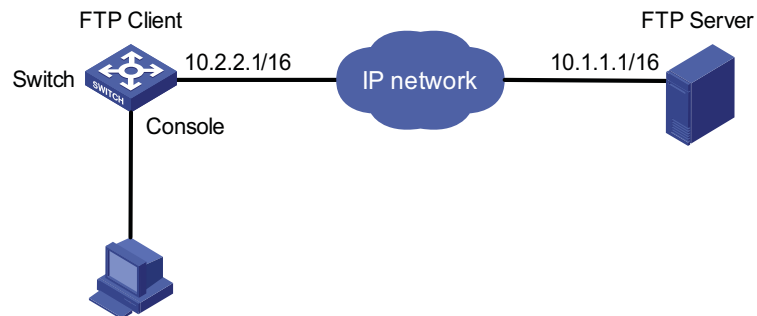
FTP Client Configuration Example

Network requirements

- Use your device as an FTP client to download a startup file from the FTP server.
- The IP address of the FTP server is 10.1.1.1/16.
- On the FTP server, an FTP user account has been created for the FTP client, with the username being **abc** and the password being **pwd**.
- The PC performs operations on the device through Console port.

Network diagram

Figure 298 Network diagram for FTPing an image file from an FTP server



Configuration procedure

Check files on your device. Remove those redundant to ensure adequate space for the startup file to be downloaded.

```

<Sysname> dir
Directory of flash:/

 0  drw-      -  Dec 07 2005 10:00:57  filename
 1  drw-      -  Jan 02 2006 14:27:51  logfile
 2  -rw-      1216  Jan 02 2006 14:28:59  config.cfg
 3  -rw-      1216  Jan 02 2006 16:27:26  backup.cfg
  
```

14605 KB total (6890 KB free)

```

<Sysname> delete /unreserved flash:/backup.cfg
  
```

Download the startup file from the server.

```

<Sysname> ftp 10.1.1.1
Trying 10.1.1.1.
Press CTRL+K to abort
Connected to 10.1.1.1
220 FTP service ready
User(10.1.1.1:(none)):abc
331 Give me your password, please
Password:
331 Password required for abc.
Password:
230 User logged in.
[ftp] binary
200 Type set to I.
[ftp] get aaa.bin bbb.bin

227 Entering Passive Mode (10.1.1.1,4,1).
125 BINARY mode data connection already open, transfer starting for aaa.bin.
....226 Transfer complete.
FTP: 5805100 byte(s) received in 19.898 second(s) 291.74Kbyte(s)/sec.
[ftp] bye
  
```

You can use the **boot-loader** command to specify the downloaded file as the main startup file for next startup. Then restart the device and the startup file of the device is updated.

```

<Sysname> boot-loader file bbb.bin main
<Sysname> reboot
  
```



CAUTION: Startup files for next startup must be saved under the root directory. You can copy or move a file to change the path of it to the root directory. For description of the corresponding command, refer to “Specifying a Boot ROM File for the Next Device Boot” on page 1040.

Configuring the FTP Server

Configuring FTP Server Operating Parameters

The FTP server uses two modes to update files when you upload files (use the **put** command) to the FTP server:

- In fast mode, the FTP server starts writing data to the Flash after file transfer completes. This protects the files intended to be overwritten on the device from being corrupted in the event that anomalies, power failure for example, occur during a file transfer.
- In normal mode, the FTP server writes data to the Flash during file transfer. This means that any anomaly, power failure for example, during file transfer might result in file corruption on the router. This mode, however, consumes less memory space than the fast mode.

Follow these steps to configure the FTP server:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the FTP server	ftp server enable	Required Disabled by default.
Configure the idle-timeout timer	ftp timeout <i>minutes</i>	Optional 30 minutes by default. In idle-timeout time, if there is no information interaction between the FTP server and client, the connection between them is terminated.
Set the file update mode in FTP	ftp update { fast normal }	Optional Normal update is used by default.

Configuring Authentication and Authorization for Accessing FTP Server

To allow an FTP user to access certain directories on the FTP server, you need to create an account for the user, authorizing access to the directories and associating the username and password with the account.

Follow these steps to configure authentication and authorization for FTP server:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a local user and enter its view	local-user <i>user-name</i>	Required No local user exists by default, and the system does not support FTP anonymous user access.

To do...	Use the command...	Remarks
Assign a password to the user	password { simple cipher } <i>password</i>	Required
Assign the FTP service to the user	service-type ftp	Required By default, the system does not support anonymous FTP access, and does not assign any service. If the FTP service is assigned, the root directory of the device is used by default.
Specify the directory an FTP user can access	work-directory <i>directory-name</i>	Optional By default, the FTP/SFTP users can access the root directory of the device.
Set the priority level of the FTP user	level <i>level</i>	Optional 0 by default To upload files to an FTP server, you need to set the FTP user level to 3.



If FTP server performs authentication, authorization and accounting (AAA) policy on FTP client, AAA related parameters should be configured on the FTP server. For more information about the **local-user**, **password**, **service-type ftp**, and **work-directory** commands and the AAA related configuration, refer to “Configuring AAA” on page 758.

FTP Server Configuration Example

Network requirements

- Use your device as an FTP server. Create a user account for an FTP user on it, setting the username to **abc** and the password to **pwd**, and setting the priority level to 3.
- The IP address of the Ethernet interface is 1.1.1.1/16.
- The PC serves as the FTP client.

Network diagram

Figure 299 Smooth upgrading using the FTP server



Configuration procedure

1 Configure Device (FTP Server)

Create an FTP user account **abc**, setting its password to **pwd**, and setting the priority level to 3.

```

<Sysname> system-view
[Sysname] local-user abc
  
```

```
[Sysname-luser-abc] password simple pwd
[Sysname-luser-abc] level 3
```

Specify abc to use FTP, and authorize its access to certain directory.

```
[Sysname-luser-abc] service-type ftp
[Sysname-luser-abc] work-directory flash:/
[Sysname-luser-abc] quit
```

Enable FTP server.

```
[Sysname] ftp server enable
[Sysname] quit
```

Check files on your device. Remove those redundant to ensure adequate space for the startup file to be uploaded.

```
<Sysname> dir
Directory of flash:/

 0  drw-      -   Dec 07 2005 10:00:57  filename
 1  drw-      -   Jan 02 2006 14:27:51  logfile
 2  -rw-     1216  Jan 02 2006 14:28:59  config.cfg
 3  -rw-     1216  Jan 02 2006 16:27:26  back.cfg
 4  drw-      -   Jan 02 2006 15:20:21  ftp
```

```
2540 KB total (2511 KB free)
```

```
<Sysname> delete /unreserved flash:/back.cfg
```

1 Configure the PC (FTP Client)

Upload the startup file to the FTP server and save it under the root directory of the FTP server.

```
c:\> ftp 1.1.1.1
Connected to 1.1.1.1.
220 FTP service ready.
User(1.1.1.1:(none)):abc
331 Password required for abc.
Password:
230 User logged in.
ftp> put aaa.bin bbb.bin
```



- When upgrading the configuration file with FTP, put the new file under the root directory.
- After you finish upgrading the Boot ROM program through FTP, you must execute the **bootrom upgrade** command to refresh the system configuration.

You can use the **boot-loader** command to specify the uploaded file as the main startup file for next startup. Then restart the device and the startup file of the device is updated.

```
<Sysname> boot-loader file bbb.bin main
<Sysname> reboot
```



CAUTION: Startup files for next startup must be saved under the root directory. You can copy or move a file to change the path of it to the root directory. For

description of the corresponding command, refer to “Specifying a Boot ROM File for the Next Device Boot” on page 1040.

Displaying and Maintaining FTP

To do...	Use the command...	Remarks
Display the configuration of the FTP client	display ftp client configuration	Available in any view
Display the configuration of the FTP server	display ftp-server	Available in any view
Display detailed information about logged-in FTP users	display ftp-user	Available in any view

81

TFTP CONFIGURATION

When configuring TFTP, go to these sections for information you are interested in:

- "TFTP Overview" on page 1001
- "Configuring the TFTP Client" on page 1002
- "Displaying and Maintaining the TFTP Client" on page 1003
- "TFTP Client Configuration Example" on page 1003

TFTP Overview

Introduction to TFTP

The Trivial File Transfer Protocol (TFTP) provides functions similar to those provided by FTP, but it is not as complex as FTP in interactive access interface and authentication. Therefore, it is more suitable where complex interaction is not needed between client and server.

TFTP uses the UDP port 69 for data transmission. For TFTP basic operation, refer to RFC 1350.

In TFTP, file transfer is initiated by the client.

- In a normal file downloading process, the client sends a read request to the TFTP server, receives data from the server, and then sends the acknowledgement to the server.
- In a normal file uploading process, the client sends a write request to the TFTP server, sends data to the server, and receives the acknowledgement from the server.

TFTP transfers files in two modes:

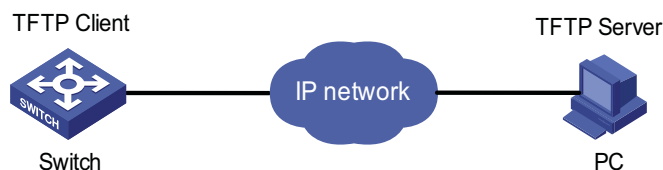
- Binary for program files
- ASCII for text files.

Implementation of TFTP



Only the TFTP client service is available with your device at present.

Figure 300 TFTP configuration diagram



Before using TFTP, the administrator needs to configure IP addresses for the TFTP client and server, and make sure that there is a route between the TFTP client and server.

Configuring the TFTP Client

When a device acts as a TFTP client, you can upload files on the device to a TFTP server and download files from the TFTP server to the local device. You can use either of the following ways to download files:

- Normal download: The device writes the obtained files to the storage device directly. In this way, the original system file will be overwritten and if file download fails (for example, due to network disconnection), the device cannot start up normally because the original system file has been deleted.
- Secure download: The device saves the obtained files to its memory and does not write them to the storage device until all user files are obtained. In this way, if file download fails (for example, due to network disconnection), the device can still start up because the original system file is not overwritten. This mode is securer but consumes more memory.

You are recommended to use the latter mode or use a filename not existing in the current directory as the target filename when downloading startup file or configuration file.

Multiple routes may exist for a TFTP client to successfully access the TFTP server. You can specify one by configuring the source address of the packets from the TFTP client to meet the requirement of the security policy of the TFTP client. You can configure the source address by configuring the source interface or source IP address. The primary IP address configured on the source interface is the source address of the transmitted packets. The source address of the transmitted packets is selected following these rules:

- If no source address of the TFTP client is specified, a device uses the IP address of the interface determined by the routing protocol as the source IP address to communicate with a TFTP server.
- If the source address is specified with the **tftp client source** or **tftp** command, this source address is adopted.
- If the source address is specified with the **tftp client source** command and then with the **tftp** command, the source address configured with the latter one is used to communicate with a TFTP server.

The source address specified with the **tftp client source** command is valid for all **tftp** connections and the source address specified with the **tftp** command is valid only for the current **tftp** connection.

Follow these steps to configure the TFTP client:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Reference an access control list (ACL) to the TFTP server	tftp-server [ipv6] acl acl-number	Optional

To do...	Use the command...	Remarks
Configure the source address of the TFTP client	tftp client source { interface <i>interface-type</i> <i>interface-number</i> ip <i>source-ip-address</i> }	Optional A device uses the source address determined by the routing protocol to communicate with the TFTP server by default.
Return to user view	quit	-
Download or upload a file in IPv4 network	tftp server-address { get put sget } <i>source-filename</i> [<i>destination-filename</i>] [source { interface <i>interface-type</i> <i>interface-number</i> ip <i>source-ip-address</i> }]	Optional
Download or upload a file in IPv6 network	tftp ipv6 <i>tftp-ipv6-server</i> [-i <i>interface-type</i> <i>interface-number</i>] { get put } <i>source-file</i> [<i>destination-file</i>]	Optional



- If no primary IP address is configured on the source interface, TFTP connection fails.
- If you use the **tftp client source** command to first configure the source interface and then the source IP address of the packets of the TFTP client, the new source IP address will overwrite the current one, and vice versa.

Displaying and Maintaining the TFTP Client

To do...	Use the command...	Remarks
Display the configuration of the TFTP client	display tftp client configuration	Available in any view

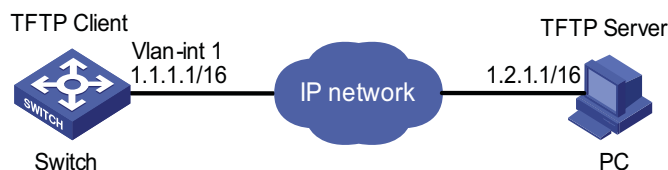
TFTP Client Configuration Example

Network requirements

- Use a PC as the TFTP server and your device as the TFTP client.
- PC uses IP address 1.2.1.1/16 and a TFTP working directory has been defined for the client.
- On your device, VLAN-interface 1 is assigned an IP address 1.1.1.1/16. Make sure that the port connected to PC belongs to the same VLAN.
- TFTP a startup file from PC for upgrading and a configuration file config.cfg to PC for backup.

Network diagram

Figure 301 Smooth upgrading using the TFTP client function



Configuration procedure

- 1 Configure PC (TFTP Server), the configuration procedure omitted.
- 2 On the PC, enable TFTP server
- 3 Configure a TFTP working directory
- 4 Configure the device (TFTP Client)



CAUTION: *If the free memory space of the device is not big enough, you should delete the existing programs before downloading new ones.*

Enter system view.

```
<Sysname> system-view
```

Assign VLAN-interface 1 an IP address 1.1.1.1/16, making sure that the port connected to PC belongs to the same VLAN.

```
[Sysname] interface vlan-interface 1
[Sysname-Vlan-interface1] ip address 1.1.1.1 255.255.0.0
[Sysname-Vlan-interface1] return
```

Download an application file aaa.bin from the TFTP server. (Before that, make sure that adequate memory is available.)

```
<Sysname> tftp 1.2.1.1 get aaa.bin bbb.bin
```

Upload a configuration file config.cfg to the TFTP server.

```
<Sysname> tftp 1.2.1.1 put config.cfg configback.cfg
```

You can use the **boot-loader** command to specify the uploaded file as the main startup file for next startup. Then restart the device and the startup file of the device is updated.

```
<Sysname> boot-loader file bbb.bin main
<Sysname> reboot
```



CAUTION: *Startup files for next startup must be saved under the root directory. You can copy or move a file to change the path of it to the root directory. For description of the corresponding command, refer to "Specifying a Boot ROM File for the Next Device Boot" on page 1040.*

INFORMATION CENTER CONFIGURATION

When configuring information center, go to these sections for information you are interested in:

- "Information Center Overview" on page 1005
- "Configuring Information Center" on page 1009
- "Displaying and Maintaining Information Center" on page 1015
- "Information Center Configuration Examples" on page 1015

Information Center Overview

Introduction to Information Center

Acting as the system information hub, information center classifies and manages system information. Together with the debugging functionality, information center offers a powerful support for network administrators and developers in monitoring network performance and diagnosing network problems.



By default, the information center is enabled. An enabled information center affects the system performance in some degree due to information classification and output. Such impact becomes more obvious in the event that there is enormous information waiting for processing.

The information center of the system has the following features:

Classification of system information

The system is available with three types of information:

- Log information
- Trap information
- Debugging information

Eight levels of system information

The information is classified into eight levels by severity and can be filtered by level. More emergent information has a smaller severity level.

Table 76 Severity description

Severity	Severity value	Description
emergencies	0	The system is unavailable.
alerts	1	Information that demands prompt reaction
critical	2	Critical information
errors	3	Error information

Table 76 Severity description

Severity	Severity value	Description
warnings	4	Warnings
notifications	5	Normal information that needs to be noticed
informational	6	Informational information to be recorded
debugging	7	Information generated during debugging

Information filtering by severity works this way: information with the severity value greater than the configured threshold is not output during the filtering.

- If the threshold is set to 0, only information with the severity being emergencies will be output;
- If the threshold is set to 7, information of all severities will be output.

Ten channels and six output destinations of system information

The system supports six information output destinations, including the console, monitor, logbuffer, loghost, trapbuffer, and SNMP.

The system supports ten channels. The channels 0 through 5 have their default channel names and are associated with six output destinations by default. Both the channel names and the associations between the channels and output destinations can be changed through commands.

Table 77 Information channels and output destinations

Information channel number	Default channel name	Default output destination
0	console	Console (Receives log, trap and debugging information)
1	monitor	Monitor terminal (Receives log, trap and debugging information, facilitating remote maintenance)
2	loghost	Log host (Receives log, trap and debugging information and information will be stored in files for future retrieval.)
3	trapbuffer	Trap buffer (Receives trap information, a buffer inside the router for recording information.)
4	logbuffer	Log buffer (Receives log information, a buffer inside the router for recording information.)
5	snmpagent	SNMP NMS (Receives trap information)
6	channel6	Not specified (Receives log, trap, and debugging information)
7	channel7	Not specified (Receives log, trap, and debugging information)
8	channel8	Not specified (Receives log, trap, and debugging information)
9	channel9	Not specified (Receives log, trap, and debugging information)



Configurations for the six output destinations function independently and take effect only after the information center is enabled.

Outputting system information by source module

The system is composed of a variety of protocol modules, module drivers, and configuration modules. The system information can be classified, filtered, and output by source module. Some source module names and descriptions are shown in Table 78.

Table 78 Source module list

Module name	Description
8021X	802.1X module
ACL	Access Control List module
ARP	Address Resolution Protocol module
BGP	Border Gateway Protocol module
CFM	Configuration File Management module
CLST	Cluster Configuration module
CMD	Command line module
default	Default setting of all modules
DEV	Device management module
DHCP	Dynamic Host Configuration Protocol module
DNS	Domain Name System module
ETH	Ethernet module
FTPS	FTP Server module
GARP	Generic Attribute Registration Protocol module
HABP	3Com Authentication Bypass Protocol module
HWCM	3Com Configuration Management MIB module
IFNET	Interface management module
IP	Internet Protocol module
ISIS	Intermediate System-to-Intermediate System intra-domain routing information exchange protocol module
LAGG	Link Aggregation module
LINE	Line module
MSDP	Multicast Source Discovery Protocol module
MSTP	Multiple Spanning Tree Protocol module
NAT	Network Address Translation module
NTP	Network Time Protocol module
PKI	Public Key Infrastructure module
OSPF	Open Shortest Path First module
POE	Power over Ethernet module
QoS	Quality of Service module
RDS	Radius module
RM	Routing Management module
RMON	Remote monitor module
SHELL	User interface module
SNMP	Simple Network Management Protocol module
SOCKET	Socket module
SSH	Secure Shell module

Table 78 Source module list

Module name	Description
SYSMIB	System MIB module
TAC	Terminal Access Controller module
TELNET	Telnet module
UDPH	UDP Helper module
VFS	Virtual File System module
VLAN	Virtual Local Area Network module
VOS	Virtual Operating System module
VRRP	Virtual Router Redundancy Protocol module
VTY	Virtual Type Terminal module

To sum up, the major task of the information center is to output the three types of information of the modules onto the ten channels in terms of the eight severity levels and according to the user's settings, and then redirect the system information from ten channels to the six output destinations.

System Information Format

System information has the following format:

```
<priority>timestamp sysname module/level/digest:content
```



- *The closing set of angel brackets < >, the space, the forward slash /, and the colon are all required in the above format.*
- *Before the <priority>, there may be a %, "#, or * followed with a space, indicating log, alarm, or debugging information respectively.*
- *This format is the standard format of system information. After the system information is sent to the log host, the displayed format depends on the tools you use to view the logs.*

Below is an example of the format of log information to be output to a log host:

```
% <188>Sep 28 15:33:46:235 2005 MyDevice SHELL/5/LOGIN: Console login from aux0
```

What follows is a detailed explanation of the fields involved:

Priority

The priority is calculated using the following formula: $\text{facility} * 8 + \text{severity}$, in which facility represents the logging facility name and can be configured when you set the log host parameters. The facility ranges from local0 to local7 (16 to 23 in decimal integers) and defaults to local7. The facility is mainly used to mark different log sources on the log host, query and filter the logs of the corresponding log source. Severity ranges from 0 to 7. Table 76 details the value and meaning associated with each severity.

Note that there is no space between the priority and timestamp fields and that the priority takes effect only when the information has been sent to the log host.

Timestamp

Timestamp records the time when system information is generated to allow users to check and identify system events.

Note that there is a space between the timestamp and sysname (host name) fields.

Sysname

Sysname is the system name of the current host. You can use the **sysname** command to modify the system name. (Refer to "System Maintaining" on page 1035 for details)

Note that there is a space between the sysname and module fields.

Module

The module field represents the name of the module that generates system information. You can enter the **info-center source ?** command in system view to view the module list.

Refer to Table 78 for module name and description.

Between "module" and "level" is a "/" .

Level (Severity)

System information can be divided into eight levels based on its severity, from 0 to 7. Refer to Table 76 for definition and description of these severity levels. Note that there is a forward slash between the levels (severity) and digest fields.

Digest

The digest field is a string of up to 32 characters, outlining the system information.

Note that there is a colon between the digest and content fields.

Content

This field provides the content of the system information.

Configuring Information Center

Information Center Configuration Task List

Complete the following tasks to configure information center:

Task	Remarks
"Setting to Output System Information to the Console" on page 1009	Optional
"Setting to Output System Information to a Monitor Terminal" on page 1011	Optional
"Setting to Output System Information to a Log Host" on page 1012	Optional
"Setting to Output System Information to the Trap Buffer" on page 1012	Optional
"Setting to Output System Information to the Log Buffer" on page 1013	Optional
"Setting to Output System Information to the SNMP NMS" on page 1013	Optional
"Configuring Synchronous Information Output" on page 1014	Optional

Setting to Output System Information to the Console

Setting to output system information to the console

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Enable information center	info-center enable	Optional Enabled by default
Name the channel with a specified channel number	info-center channel <i>channel-number name</i> <i>channel-name</i>	Optional Refer to Table 77 for default channel names.
Configure the channel through which system information can be output to the console	info-center console channel { <i>channel-number</i> <i>channel-name</i> }	Optional System information is output to the console by default, with channel 0 as the default channel.
Configure the output rules of system information	info-center source { <i>module-name</i> default } channel { <i>channel-number</i> <i>channel-name</i> } [debug { <i>level severity</i> state state } * log { <i>level severity</i> state state } * trap { <i>level severity</i> state state } *] *	Optional Refer to Table 79 for the default output rules of system information.
Configure the format of the time stamp	info-center timestamp { debugging log trap } { boot date none }	Optional The time stamp format for log, trap and debugging information is date by default.

Table 79 Default output rules for different output destinations

Output destination	Modules allowed	LOG		TRAP		DEBUG	
		Enabled/disabled	Severity	Enabled/disabled	Severity	Enabled/disabled	Severity
Console	default (all modules)	Enabled	warnings	Enabled	debugging	Enabled	debugging
Monitoring terminal	default (all modules)	Enabled	warnings	Enabled	debugging	Enabled	debugging
Log host	default (all modules)	Enabled	informational	Enabled	debugging	Disabled	debugging
Trap buffer	default (all modules)	Disabled	informational	Enabled	warnings	Disabled	debugging
Log buffer	default (all modules)	Enabled	warnings	Disabled	debugging	Disabled	debugging
SNMP NMS	default (all modules)	Disabled	debugging	Enabled	warnings	Disabled	debugging

Enabling the display of system information on the console

After setting to output system information to the console, you need to enable the associated display function to display the output information on the console.

Follow these steps in user view to enable the display of system information on the console:

To do...	Use the command...	Remarks
Enable the monitoring of system information on the console	terminal monitor	Optional Enabled on the console and disabled on the monitoring terminal by default.
Enable the display of debugging information on the console	terminal debugging	Required Disabled by default

Setting to Output System Information to a Monitor Terminal

System information can also be output to a monitor terminal, which is a user terminal that has login connections through the AUX, or VTY user interface.

Setting to output system information to a monitor terminal

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable information center	info-center enable	Optional Enabled by default
Name the channel with a specified channel number	info-center channel <i>channel-number</i> name <i>channel-name</i>	Optional Refer to Table 77 for default channel names.
Configure the channel through which system information can be output to a monitor terminal	info-center monitor channel { <i>channel-number</i> <i>channel-name</i> }	Optional System information is output to the monitor terminal by default with channel 1 as the default channel.
Configure the output rules of the system information	info-center source { <i>module-name</i> default } channel { <i>channel-number</i> <i>channel-name</i> } [debug { <i>level severity</i> state state } * log { <i>level severity</i> state state } * trap { <i>level severity</i> state state } *] *	Optional Refer to Table 79 for the default output rules of the system information.
Configure the format of the time stamp	info-center timestamp { debugging log trap } { boot date none }	Optional By default, the time stamp format for log, trap and debugging information is date .

Enabling the display of system information on a monitor terminal

After setting to output system information to a monitor terminal, you need to enable the associated display function in order to display the output information on the monitor terminal.

Follow these steps to enable the display of system information on a monitor terminal:

To do...	Use the command...	Remarks
Enable the monitoring of system information on a monitor terminal	terminal monitor	Required Enabled on the console disabled on the monitoring terminal by default.

To do...	Use the command...	Remarks
Enable the display of debugging information on a monitor terminal	terminal debugging	Required Disabled by default
Enable the display of log information on a monitor terminal	terminal logging	Optional Enabled by default
Enable the display of trap information on a monitor terminal	terminal trapping	Optional Enabled by default

Setting to Output System Information to a Log Host

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable information center	info-center enable	Optional Enabled by default
Name the channel with a specified channel number	info-center channel <i>channel-number name</i> <i>channel-name</i>	Optional Refer to Table 77 for default channel names.
Specify a log host and configure the parameters when system information is output to the log host	info-center loghost <i>host-ip</i> [channel { <i>channel-number</i> <i>channel-name</i> }] facility <i>local-number</i>] *	Required By default, the system does not output information to a log host. If you specify to output system information to a log host, the system uses channel 2 (loghost) by default.
Configure the source interface through which log information can be output to a log host	info-center loghost source <i>interface-type</i> <i>interface-number</i>	Optional No source interface is configured by default, and the system selects an interface as the source interface.
Configure the output rules of the system information	info-center source { <i>module-name</i> default } channel { <i>channel-number</i> <i>channel-name</i> } [debug { level <i>severity</i> state <i>state</i> } * log { level <i>severity</i> state <i>state</i> } * trap { level <i>severity</i> state <i>state</i> } *] *	Optional Refer to Table 79 for the output rules of the system information.
Configure the format of the time stamp for log information	info-center timestamp loghost { date no-year-date none }	Optional date by default.

Setting to Output System Information to the Trap Buffer

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable information center	info-center enable	Optional Enabled by default
Name the channel with a specified channel number	info-center channel <i>channel-number name</i> <i>channel-name</i>	Optional Refer to Table 77 for default channel names.

To do...	Use the command...	Remarks
Configure the channel through which system information can be output to the trap buffer and specify the buffer size	info-center trapbuffer [channel { <i>channel-number</i> <i>channel-name</i> } size <i>buffersize</i>] *	Optional System information is output to the trap buffer by default with a default channel of channel 3 (known as trapbuffer) and a default buffer size of 256.
Configure the output rules of the system information	info-center source { <i>module-name</i> default } channel { <i>channel-number</i> <i>channel-name</i> } [debug { <i>level severity</i> state state } * log { <i>level severity</i> state state } * trap { <i>level severity</i> state state } *] *	Optional Refer to Table 79 for the output rules of the system information.
Configure the format of the time stamp	info-center timestamp { debugging log trap } { boot date none }	Optional The time stamp format for log, trap and debugging information is date by default.

Setting to Output System Information to the Log Buffer

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable information center	info-center enable	Optional Enabled by default.
Name the channel with a specified channel number	info-center channel <i>channel-number</i> name <i>channel-name</i>	Optional Refer to Table 77 for default channel names.
Configure the channel through which system information can be output to the log buffer and specify the buffer size	info-center logbuffer [channel { <i>channel-number</i> <i>channel-name</i> } size <i>buffersize</i>] *	Optional System information is output to the log buffer by default with a default channel of channel 4 (known as logbuffer) and a default buffer size of 512.
Configure the output rules of the system information	info-center source { <i>module-name</i> default } channel { <i>channel-number</i> <i>channel-name</i> } [debug { <i>level severity</i> state state } * log { <i>level severity</i> state state } * trap { <i>level severity</i> state state } *] *	Optional Refer to Table 79 for the output rules of the system information.
Configure the format of the timestamp	info-center timestamp { debugging log trap } { boot date none }	Optional The time stamp format for log, trap and debugging information is date by default.

Setting to Output System Information to the SNMP NMS

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Enable information center	info-center enable	Optional Enabled by default
Name the channel with a specified channel number	info-center channel <i>channel-number name</i> <i>channel-name</i>	Optional Refer to Table 77 for default channel names.
Configure the channel through which system information can be output to the SNMP NMS	info-center snmp channel { <i>channel-number</i> <i>channel-name</i> }	Optional System information is output to the SNMP NMS by default with channel 5 (known as snmpagent) as the default channel.
Configure the output rules of the system information	info-center source { <i>module-name</i> default } channel { <i>channel-number</i> <i>channel-name</i> } [debug { level <i>severity</i> state <i>state</i> } * log { level <i>severity</i> state <i>state</i> } * trap { level <i>severity</i> state <i>state</i> } *] *	Optional Refer to Table 79 for the output rules of the system information.
Configure the format of the timestamp	info-center timestamp { debugging log trap } { boot date none }	Optional The time stamp format for log, trap and debugging information is date by default.



To ensure that system information can be output to the SNMP NMS, you need to make the necessary configurations on the SNMP agent and the NMS. For detailed information on SNMP, refer to the “SNMP Configuration” on page 931.

Configuring Synchronous Information Output

Synchronous information output refers to the feature that if the user’s input is interrupted by system output such as log, trap, or debugging information, then after the completion of system output the system will display a command line prompt (in command editing mode a prompt, or a [Y/N] string in interaction mode) and your input so far.

This command is used in the case that your input is interrupted by a large amount of system output. With this feature enabled, you can continue your operations from where you were stopped.

Follow these steps to enable synchronous information output:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable synchronous information output	info-center synchronous	Required Disabled by default



- If you do not input any information following the current command line prompt, the system does not display any command line prompt after system information output.

- In the interaction mode, you are prompted for some information input. If the input is interrupted by system output, no system prompt will be made, rather only your input will be displayed in a new line.

Displaying and Maintaining Information Center

To do...	Use the command...	Remarks
Display channel information for a specified channel	display channel [<i>channel-number</i> <i>channel-name</i>]	Available in any view
Display the configurations on each output destination	display info-center	Available in any view
Display the state of the log buffer and the log information recorded	display logbuffer [<i>level severity</i> <i>size buffersize</i>] * [{ begin exclude include } <i>text</i>]	Available in any view
Display a summary of the log buffer	display logbuffer summary [<i>level severity</i>]	Available in any view
Display the state of the trap buffer and the trap information recorded	display trapbuffer [<i>size buffersize</i>]	Available in any view
Reset the log buffer	reset logbuffer	Available in user view
Reset the trap buffer	reset trapbuffer	Available in user view

Information Center Configuration Examples

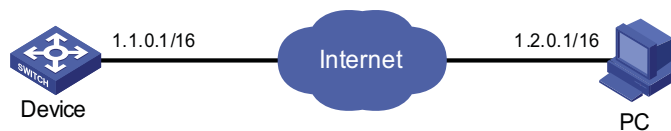
Outputting Log Information to a Unix Log Host

Network requirements

- Send log information to a Unix log host with an IP address of 1.2.0.1/16;
- Log information with severity higher than informational will be output to the log host;
- The source modules are ARP and IP.

Network diagram

Figure 302 Network diagram for outputting log information to a Unix log host



Configuration procedure

Before the configuration, make sure that there is a route between Device and PC.

1 Configuring the device

```
# Enable information center.
```

```
<Sysname> system-view
[Sysname] info-center enable
```

Specify the host with IP address 1.2.0.1/16 as the log host, use channel **loghost** to output log information (optional, **loghost** by default), and specify **local4** as the logging facility.

```
[Sysname] info-center loghost 1.2.0.1 channel loghost facility local4
```

Disable the output of log, trap, and debugging information of all modules on the channel **loghost**.

```
[Sysname] info-center source default channel loghost debug state off log state off trap state off
```



CAUTION: As the default system configurations for different channels are different, you need to disable the output of log, trap, and debugging information of all modules on the specified channel (loghost in this example) first and then configure the output rule as needed so that unnecessary information will not be output.

Configure the information output rule: allow log information of ARP and IP modules with severity equal to or higher than **informational** to be output to the log host.

```
[Sysname] info-center source arp channel loghost log level informational state on
[Sysname] info-center source ip channel loghost log level informational state on
```

2 Configuring the log host

The following configurations were performed on SunOS 4.0 which has similar configurations to the Unix operating systems implemented by other vendors.

Step 1: Issue the following commands as a root user.

```
# mkdir /var/log/MyDevice
# touch /var/log/MyDevice/information
```

Step 2: Edit the file /etc/syslog.conf as a root user and add the following selector/action pair.

```
# MyDevice configuration messages
local4.info /var/log/MyDevice/information
```



Be aware of the following issues while editing the /etc/syslog.conf file

- Comments must be on a separate line and must begin with the # sign.
- The selector/action pair must be separated with a tab key, rather than a space.
- No redundant spaces are allowed in the file name.
- The device name and the accepted severity of log information specified by the /etc/syslog.conf file must be identical to those configured on the device using the **info-center loghost** or **info-center source** command; otherwise the log information may not be output properly to the log host.

Step three: After the log file information has been created and the configuration file /etc/syslog.conf has been modified, ensure that the configuration file /etc/syslog.conf is reread:

```
# ps -ae | grep syslogd
147
# kill -HUP 147
```

After the above configurations, the system will be able to keep log information in the related file.

Outputting Log Information to a Linux Log Host

Network requirements

- Send log information to a Linux log host with an IP address of 1.2.0.1/16;
- Log information with severity higher than informational will be output to the log host;
- All modules can output log information.

Network diagram

Figure 303 Network diagram for outputting log information to a Linux log host



Configuration procedure

Before configuration, make sure that there is a route between Device and PC.

1 Configuring the device

Enable information center.

```
<Sysname> system-view
[Sysname] info-center enable
```

Specify the host with IP address 1.2.0.1/16 as the log host, use channel **loghost** to output log information (optional, **loghost** by default), and specify **local5** as the logging facility.

```
[Sysname] info-center loghost 1.2.0.1 channel loghost facility local5
```

Disable the output of log, trap, and debugging information of all modules on the channel **loghost**.

```
[Sysname] info-center source default channel loghost debug state off log state off trap state off
```



CAUTION: As the default system configurations for different channels are different, you need to disable the output of log, trap, and debugging information of all modules on the specified channel (*loghost* in this example) first and then configure the output rule as needed so that unnecessary information will not be output.

Configure the information output rule: allow log information of all modules with severity equal to or higher than **informational** to be output to the log host.

```
[Sysname] info-center source default channel loghost log level informational state on
```

2 Configuring the log host

Step 1: Issue the following commands as a root user.

```
# mkdir /var/log/MyDevice
# touch /var/log/MyDevice/information
```

Step 2: Edit the file `/etc/syslog.conf` as a root user and add the following selector/action pair.

```
# MyDevice configuration messages
local7.info /var/log/MyDevice/information
```



Be aware of the following issues while editing the `/etc/syslog.conf` file

- Comments must be on a separate line and must begin with the `#` sign.
- The selector/action pair must be separated with a tab key, rather than a space.
- No redundant spaces are allowed in the file name.
- The device name and the accepted severity of the log information specified by the `/etc/syslog.conf` file must be identical to those configured on the device using the **info-center loghost** or **info-center source** command; otherwise the log information may not be output properly to the log host.

Step 3: After the log file information has been created and the `/etc/syslog.conf` file has been modified, issue the following commands to display the process ID of **syslogd**, terminate a **syslogd** process, and restart **syslogd** using the `-r` option.

```
# ps -ae | grep syslogd
147
# kill -9 147
# syslogd -r &
```



Ensure that the `syslogd` process is started with the `-r` option on a Linux log host.

After the above configurations, the system will be able to keep log information in the related file.

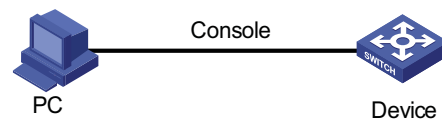
Outputting Log Information to the Console

Network requirements

- Log information with a severity higher than informational will be output to the console;
- The source modules are ARP and IP.

Network diagram

Figure 304 Network diagram for sending log information to the console



Configuration procedure

```
# Enable information center.
```

```
<Sysname> system-view
[Sysname] info-center enable
```

Use channel **console** to output log information to the console (optional, **console** by default).

```
[Sysname] info-center console channel console
```

Disable the output of log, trap, and debugging information of all modules on the channel **console**.

```
[Sysname] info-center source default channel console debug state off log state off trap state off
```



CAUTION: As the default system configurations for different channels are different, you need to disable the output of log, trap, and debugging information of all modules on the specified channel (**console** in this example) first and then configure the output rule as needed so that unnecessary information will not be output.

Configure the information output rule: allow log information of ARP and IP modules with severity equal to or higher than **informational** to be output to the console.

```
[Sysname] info-center source arp channel console log level informational state on
[Sysname] info-center source ip channel console log level informational state on
[Sysname] quit
```

Enable the display of log information on a monitor terminal.

```
<Sysname> terminal monitor
% Current terminal monitor is on
<Sysname> terminal logging
% Current terminal logging is on
```

After the above configuration takes effect, if the specified module generates log information, the information center automatically sends the log information to the console and displays it on the console.

BASIC CONFIGURATIONS

While performing basic configurations of the system, go to these sections for information you are interested in:

- “Basic Configurations” on page 1021
- “CLI Features” on page 1027

Basic Configurations

This section covers the following topics:

- “Entering/Exiting System View” on page 1021
- “Configuring the Device Name” on page 1021
- “Configuring the System Clock” on page 1021
- “Configuring a Banner” on page 1023
- “Configuring CLI Hotkeys” on page 1025
- “Configuring User Levels and Command Levels” on page 1026
- “Displaying and Maintaining Basic Configurations” on page 1027

Entering/Exiting System View

Follow these steps to enter/exit system view:

To do...	Use the command...	Remarks
Enter system view from user view	system-view	-
Return to user view from system view	quit	-



*With the **quit** command, you can return to the previous view. You can execute the **return** command or press the hot key <Ctrl+Z> to return to user view.*

Configuring the Device Name

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the device name	sysname <i>sysname</i>	Optional The device name is 3Com by default.

Configuring the System Clock

Configuring the system clock

Follow these steps to configure the system clock:

To do...	Use the command...	Remarks
Set time and date	clock datetime <i>time date</i>	Optional
Set the time zone	clock timezone <i>zone-name</i> { add minus } <i>zone-offset</i>	Available in user view.
Set a summer time scheme	clock summer-time <i>zone-name</i> one-off <i>start-time start-date end-time end-date add-time</i> clock summer-time <i>zone-name</i> repeating <i>start-time start-date end-time end-date add-time</i>	

Displaying the system clock

The system clock is displayed by system time stamp, which is the same as that displayed by the **display clock** command. The system clock is decided by the commands **clock datetime**, **clock timezone** and **clock summer-time**. If these three commands are not configured, the **display clock** command displays the original system clock. If you combine these three commands in different ways, the system clock is displayed in the ways shown in Table 80. The meanings of the parameters in the configuration column are as follows:

- 1 indicates date-time has been configured with the **clock datetime**.
- 2 indicates time-zone has been configured with the **clock timezone** command and the offset time is *zone-offset*.
- 3 indicates summer time has been configured with the **clock summer-time** command and the offset time is *summer-offset*.
- [1] indicates the **clock datetime** command is an optional configuration.
- The default system clock is 2005/1/1 1:00:00 in the example.

Table 80 Relationship between the configuration and display of the system clock

Configuration	System clock displayed by the display clock command	Example
1	<i>date-time</i>	Configure: clock datetime 1:00 2007/1/1 Display: 01:00:00 UTC Mon 01/01/2007
2	The original system clock $\text{DC} \rightarrow \pm \text{zone-offset}$	Configure: clock timezone zone-time add 1 Display: 02:00:00 zone-time Sat 01/01/2005
1 and 2	<i>date-time</i> $\text{DC} \rightarrow \pm \text{zone-offset}$	Configure: clock datetime 2:00 2007/2/2 and clock timezone zone-time add 1 Display: 03:00:00 zone-time Fri 02/02/2007
[1], 2 and 1	<i>date-time</i>	Configure: clock timezone zone-time add 1 and clock datetime 3:00 2007/3/3 Display: 03:00:00 zone-time Sat 03/03/2007
3	If the original system clock is not in the summer time range, the original system clock is displayed. If the original system clock is in the summer time range, the original system clock + <i>summer-offset</i> is displayed.	Configure: clock summer-time ss one-off 1:00 2006/1/1 1:00 2006/8/8 2 Display: 01:00:00 UTC Sat 01/01/2005 Configure: clock summer-time ss one-off 00:30 2005/1/1 1:00 2005/8/8 2 Display: 03:00:00 ss Sat 01/01/2005

Table 80 Relationship between the configuration and display of the system clock

Configuration	System clock displayed by the display clock command	Example
1 and 3	If <i>date-time</i> is not in the summer time range, <i>date-time</i> is displayed. If <i>date-time</i> is in the summer time range, " <i>date-time</i> " + " <i>summer-offset</i> " is displayed.	Configure: clock datetime 1:00 2007/1/1 and clock summer-time ss one-off 1:00 2006/1/1 1:00 2006/8/8 2 Display: 01:00:00 UTC Mon 01/01/2007 Configure: clock datetime 8:00 2007/1/1 and clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 Display: 10:00:00 ss Mon 01/01/2007
[1], 3 and 1	If <i>date-time</i> is not in the summer time range, <i>date-time</i> is displayed. <i>date-time</i> is in the summer time range: If the value of " <i>date-time</i> " - " <i>summer-offset</i> " is not in the summer-time range, " <i>date-time</i> " - " <i>summer-offset</i> " is displayed; If the value of " <i>date-time</i> " - " <i>summer-offset</i> " is in the summer-time range, <i>date-time</i> is displayed.	Configure: clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 and clock datetime 1:00 2008/1/1 Display: 01:00:00 UTC Tue 01/01/2008 Configure: clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 and clock datetime 1:30 2007/1/1 Display: 23:30:00 UTC Sun 12/31/2006 Configure: clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 and clock datetime 3:00 2007/1/1 Display: 03:00:00 ss Mon 01/01/2007
2 and 3 or 3 and 2	If the value of the original system clock $\text{ĐÇ}\rightarrow\pm$ " <i>zone-offset</i> " is not in the summer-time range, the original system clock $\text{ĐÇ}\rightarrow\pm$ " <i>zone-offset</i> " is displayed. If the value of the original system clock $\text{ĐÇ}\rightarrow\pm$ " <i>zone-offset</i> " is in the summer-time range, the original system clock $\text{ĐÇ}\rightarrow\pm$ " <i>zone-offset</i> " + " <i>summer-offset</i> " is displayed.	Configure: clock timezone zone-time add 1 and clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 Display: 02:00:00 zone-time Sat 01/01/2005 Configure: clock timezone zone-time add 1 and clock summer-time ss one-off 1:00 2005/1/1 1:00 2005/8/8 2 Display: 04:00:00 ss Sat 01/01/2005 Configure: clock datetime 1:00 2007/1/1, clock timezone zone-time add 1 and clock summer-time ss one-off 1:00 2008/1/1 1:00 2008/8/8 2 Display: 02:00:00 zone-time Mon 01/01/2007
1, 2 and 3 or 1, 3 and 2	If the value of " <i>date-time</i> " $\text{ĐÇ}\rightarrow\pm$ " <i>zone-offset</i> " is not in the summer-time range, " <i>date-time</i> " $\text{ĐÇ}\rightarrow\pm$ " <i>zone-offset</i> " is displayed. If the value of " <i>date-time</i> " $\text{ĐÇ}\rightarrow\pm$ " <i>zone-offset</i> " is in the summer-time range, " <i>date-time</i> " $\text{ĐÇ}\rightarrow\pm$ " <i>zone-offset</i> " + " <i>summer-offset</i> " is displayed.	Configure: clock datetime 1:00 2007/1/1, clock timezone zone-time add 1 and clock summer-time ss one-off 1:00 2007/1/1 1:00 2007/8/8 2 Display: 04:00:00 ss Mon 01/01/2007 Configure: clock timezone zone-time add 1, clock summer-time ss one-off 1:00 2008/1/1 1:00 2008/8/8 2 and clock datetime 1:00 2007/1/1 Display: 01:00:00 zone-time Mon 01/01/2007
[1], 2, 3 and 1 or [1], 3, 2 and 1	If <i>date-time</i> is not in the summer time range, <i>date-time</i> is displayed. <i>date-time</i> is in the summer time range: If the value of " <i>date-time</i> " - " <i>summer-offset</i> " is not in the summer-time range, " <i>date-time</i> " - " <i>summer-offset</i> " is displayed; If the value of " <i>date-time</i> " - " <i>summer-offset</i> " is in the summer-time range, <i>date-time</i> is displayed.	Configure: clock timezone zone-time add 1, clock summer-time ss one-off 1:00 2008/1/1 1:00 2008/8/8 2 and clock datetime 1:30 2008/1/1 Display: 23:30:00 zone-time Mon 12/31/2007 Configure: clock timezone zone-time add 1, clock summer-time ss one-off 1:00 2008/1/1 1:00 2008/8/8 2 and clock datetime 3:00 2008/1/1 Display: 03:00:00 ss Tue 01/01/2008

Configuring a Banner Introduction to banners

Banners are prompt information displayed by the system when users are connected to the device, perform login authentication, and start interactive configuration. The administrator can set corresponding banners as needed.

At present, the system supports the following five kinds of welcome information.

- **shell** banner, also called session banner, displayed when a non Modem user enters user view.
- **incoming** banner, also called user interface banner, displayed when a user interface is activated by a Modem user.
- **login** banner, welcome information at login authentications, displayed when password and scheme authentications are configured.
- **motd** banner, welcome information displayed before authentication.
- **legal** banner, also called authorization information. The system displays some copyright or authorization information, and then displays the **legal** banner before a user logs in, waiting for the user to confirm whether to continue the authentication or login. If entering Y or pressing the **Enter** key, the user enters the authentication or login process; if entering N, the user quits the authentication or login process. Y and N are case insensitive.

Configuring a banner

When you configure a banner, the system supports two input modes. One is to input all the banner information right after the command keywords. The start and end characters of the input text must be the same but are not part of the banner information. In this case, the input text, together with the command keywords, cannot exceed 510 characters. The other is to input all the banner information in multiple lines by pressing the **Enter** key. In this case, up to 2000 characters can be input.

The latter input mode can be achieved in the following three ways:

- Press the **Enter** key directly after the command keywords, and end the setting with the % character. The **Enter** and % characters are not part of the banner information.
- Input a character after the command keywords at the first line, and then press the **Enter** key. End the setting with the character input at the first line. The character at the first line and the end character are not part of the banner information.
- Input multiple characters after the command keywords at the first line (with the first and last characters being different), then press the **Enter** key. End the setting with the first character at the first line. The first character at the first line and the end character are not part of the banner information.

Follow these steps to configure a banner:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the banner to be displayed at login	header incoming <i>text</i>	Optional
Configure the banner to be displayed at login authentication	header login <i>text</i>	Optional
Configure the authorization information before login	header legal <i>text</i>	Optional
Configure the banner to be displayed when a user enters user view	header shell <i>text</i>	Optional

To do...	Use the command...	Remarks
Configure the banner to be displayed before login	header motd <i>text</i>	Optional

Configuring CLI Hotkeys

Follow these steps to configure CLI hotkeys:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure CLI hotkeys	hotkey { CTRL_G CTRL_L CTRL_O CTRL_T CTRL_U } <i>command</i>	Optional The <Ctrl+G>, <Ctrl+L> and <Ctrl+O> hotkeys are specified with command lines by default.
Display hotkeys	display hotkey	Available in any view. Refer to Table 81 for hotkeys reserved by the system.



By default, the <Ctrl+G>, <Ctrl+L> and <Ctrl+O> hotkeys are configured with command line and the <Ctrl+T> and <Ctrl+U> commands are NULL.

- <Ctrl+G> corresponds to the **display current-configuration** command.
- <Ctrl+L> corresponds to the **display ip routing-table** command.
- <Ctrl+O> corresponds to the **undo debugging all** command.

Table 81 Hotkeys reserved by the system

Hotkey	Function
<Ctrl+A>	Moves the cursor to the beginning of the current line.
<Ctrl+B>	Moves the cursor one character to the left.
<Ctrl+C>	Stops performing a command.
<Ctrl+D>	Deletes the character at the current cursor position.
<Ctrl+E>	Moves the cursor to the end of the current line.
<Ctrl+F>	Moves the cursor one character to the right.
<Ctrl+H>	Deletes the character to the left of the cursor.
<Ctrl+K>	Terminates an outgoing connection.
<Ctrl+N>	Displays the next command in the history command buffer.
<Ctrl+P>	Displays the previous command in the history command buffer.
<Ctrl+R>	Redisplays the current line information.
<Ctrl+V>	Pastes the content in the clipboard.
<Ctrl+W>	Deletes all the characters in a continuous string to the left of the cursor.
<Ctrl+X>	Deletes all the characters to the left of the cursor.
<Ctrl+Y>	Deletes all the characters to the right of the cursor.
<Ctrl+Z>	Exits to user view.
<Ctrl+]>	Terminates an incoming connection or a redirect connection.
<Esc+B>	Moves the cursor to the leading character of the continuous string to the left.
<Esc+D>	Deletes all the characters of the continuous string at the current cursor position and to the right of the cursor.
<Esc+F>	Moves the cursor to the front of the next continuous string to the right.
<Esc+N>	Moves the cursor down by one line (available before you press the Enter key)
<Esc+P>	Moves the cursor up by one line (available before you press the Enter key)

Table 81 Hotkeys reserved by the system

Hotkey	Function
<Esc+<>	Specifies the cursor as the beginning of the clipboard.
<Esc+>>	Specifies the cursor as the ending of the clipboard.



These hotkeys are defined by the device. When you interact with the device from terminal software, these keys may be defined to perform other operations. If so, the definition of the terminal software will dominate.

Configuring User Levels and Command Levels

All the commands are defaulted to different views and categorized into four levels: visit, monitor, system, and manage, identified respectively by 0 through 3. If you want to acquire a higher privilege, you must switch to a higher user level, and it requires password to do so for the security's sake.

The following table describes the default level of the commands.

Table 82 Default command levels

Level	Privilege	Command
0	Visit	ping, tracert, telnet
1	Monitor	refresh, reset, send
2	System	All configuration commands except for those at manage level
3	Manage	FTP, TFTP, XMODEM, and file system operation commands

Follow these steps to configure user level and command level:

To do...	Use the command...	Remarks
Switch the user level	super [<i>level</i>]	Optional Available in user view.
Enter system view	system-view	-
Configure the password for switching the user level	super password [<i>level</i> <i>user-level</i>] { simple cipher } <i>password</i>	Optional By default, no password is configured.
Configure the command level in system view	command-privilege level <i>level</i> view <i>view command</i>	Optional



The commands available depend on your user level when you log onto a device. For example, if your user level is 3 and the command level of VTY 0 interface is 1, you can use commands below level 3 (inclusive).



CAUTION:

- When you configure the password for switching user level with the **super password** command, the user level is defaulted to 3 if no user level is specified.
- You can switch to a lower user level unconditionally. To switch to a higher user level, however, you need to enter the password needed (The password can be set with the **super password** command.). If the entered password is incorrect or no password is configured, the switch fails. Therefore, before switching to a higher user level, you should configure the password needed.

- You are recommended to use the default user level; otherwise the change of user level may bring inconvenience to your maintenance and operation.

Displaying and Maintaining Basic Configurations

To do...	Use the command...	Remarks
Display information on system version	display version	-
Display information on the system clock	display clock	-
Display information on terminal users	display users [all]	-
Display the configurations saved in the storage device	display saved-configuration [by-linenum]	-
Display the current validated configurations	display current-configuration [[configuration [configuration] controller interface [interface-type] [interface-number]] [by-linenum] [{ begin include exclude } text]]	-
Display the valid configuration under current view	display this [by-linenum]	-
Display clipboard information	display clipboard	-
Display and save statistics of each module's running status	display diagnostic-information	-

During daily maintenance or when the system is operating abnormally, you need to view each module's running status to find the problem. Therefore, you are required to execute the corresponding **display** commands one by one. To collect more information one time, you can execute the **display diagnostic-information** command in any view to display statistics of each module's running status. The execution of the **display diagnostic-information** command has the same effect as that of the commands **display clock**, **display version**, **display device**, and **display current-configuration**.



- For the detailed description of the **display users** command, refer to "Controlling Login Users" on page 75.
- The **display** commands discussed above are for the global configuration. Refer to the corresponding section for the **display** command for specific protocol and interface.

CLI Features

This section covers the following topics:

- "Introduction to CLI" on page 1028
- "Online Help with Command Lines" on page 1028
- "Synchronous Information Output" on page 1029
- "undo Form of a Command" on page 1029
- "Edit Features" on page 1029
- "CLI Display" on page 1030
- "Saving History Commands" on page 1031
- "Command Line Error Information" on page 1031

Introduction to CLI CLI is an interaction interface between devices and users. Through CLI, you can configure your devices by entering commands and view the output information and verify your configurations, thus facilitating your configuration and management of your devices.

CLI provides the following features for you to configure and manage your devices:

- Hierarchical command protection where you can only execute the commands at your own or lower levels. Refer to “Configuring User Levels and Command Levels” on page 1026 for details.
- Easy access to on-line help by entering “?”
- Abundant debugging information for fault diagnosis.
- Saving and executing commands that have been executed.
- Fuzzy match for convenience of input. You only need to input the characters that can uniquely identify a keyword to recognize and execute the keyword. For example, for the keyword **Ethernet**, you only need to input **eth** when you execute a command with this keyword.

Online Help with Command Lines

The following are the types of online help available with the CLI:

- Full help
- Fuzzy help

To obtain the desired help information, you can:

- 1 Enter <?> in any view to access all the commands in this view and brief description about them as well.

```
<Sysname> ?
User view commands:
  backup          Backup next startup-configuration file to TFTP server
  boot-loader     Set boot loader
  bootrom         Update/read/backup/restore bootrom
  cd              Change current directory
  clock           Specify the system clock
  cluster         Run cluster command
  copy            Copy from one file to another
  debugging       Enable system debugging functions
  delete          Delete a file
  dir             List files on a file system
  display         Display current system information
.....omitted.....
```

- 2 Enter a command and a <?> separated by a space. If <?> is at the position of a keyword, all the keywords are given with a brief description.

```
<Sysname> terminal ?
  debugging Send debug information to terminal
  logging   Send log information to terminal
  monitor   Send information output to current terminal
  trapping  Send trap information to terminal
```

- 3 Enter a command and a <?> separated by a space. If <?> is at the position of a parameter, the description about this parameter is given.

```
<Sysname> system-view
[Sysname] interface vlan-interface
<1-4094> VLAN interface number
```



```
[Sysname] interface vlan-interface 1 ?
<cr>
[Sysname] interface vlan-interface 1
```

Where, <cr> indicates that there is no parameter at this position. The command is then repeated in the next command line and executed if you press <Enter>.

- 4 Enter a character string followed by a <?>. All the commands starting with this string are displayed.

```
<Sysname> c?
  cd
  clock
  copy
```

- 5 Enter a command followed by a character string and a <?>. All the keywords starting with this string are listed.

```
<Sysname> display ver?
  version
```

- 6 Press <Tab> after entering the first several letters of a keyword to display the complete keyword, provided these letters can uniquely identify the keyword in this command.

Synchronous Information Output

Synchronous information output refers to the feature that if the user's input is interrupted by system output, then after the completion of system output the system will display a command line prompt and your input so far, and you can continue your operations from where you were stopped.

You can use the **info-center synchronous** command to enable synchronous information output. For the detailed description of this function, refer to "Configuring Synchronous Information Output" on page 1014.

undo Form of a Command

Adding the keyword **undo** can form an **undo** command. Almost every configuration command has the **undo** form. **undo** commands are generally used to restore the system default, disable a function or cancel a configuration. For example, the **info-center enable** command is used to enable the information center, while the **undo info-center enable** command is used to disable the information center. (By default, the information center is enabled.)

Edit Features

The CLI provides the basic command edit functions and supports multi-line editing. The maximum length of each command is 256 characters. Table 83 lists these functions.

Table 83 Edit functions

Key	Function
Common keys	If the editing buffer is not full, insert the character at the position of the cursor and move the cursor to the right.
<Backspace> key	Deletes the character to the left of the cursor and move the cursor back one character.
Left-arrow key or <Ctrl+B>	The cursor moves one character space to the left.
Right-arrow key or <Ctrl+F>	The cursor moves one character space to the right.
Up-arrow key or <Ctrl+P>	Displays history commands
Down-arrow key or <Ctrl+N>	

Table 83 Edit functions

Key	Function
<Tab> key	Pressing <Tab> after entering part of a keyword enables the fuzzy help function. If finding a unique match, the system substitutes the complete keyword for the incomplete one and displays it in the next line. If there are several matches or no match at all, the system does not modify the incomplete keyword and displays it again in the next line.



When editing command line, you can use other shortcut keys (For details, see Table 81) besides the shortcut keys defined in Table 83, or you can define shortcut keys by yourself. (For details, see “Configuring CLI Hotkeys” on page 1025.)

CLI Display **Filtering the output information**

The device provides the function to filter the output information. You can specify a regular expression to locate and search information you need.

The regular expression is a string of 1 to 256 characters, case sensitive, and space allowed. It supports multiple mapping rules:

- **begin**: Displays the configuration beginning with the specified regular expression.
- **exclude**: Displays the configuration excluding the specified regular expression.
- **include**: Displays the configuration including the specified regular expression.

The regular expression also supports special characters as shown in Table 84.

Table 84 Special characters in a regular expression

Character	Meaning	Remarks
^	Starting sign, the string following it appears only at the beginning of a line.	Regular expression “^user” matches a string begins with “user”, not “Auser”.
\$	Ending sign, the string before it appears only at the end of a line.	Regular expression “user\$” matches a string ends with “user”, not “userA”.
.	Full stop, a wildcard used in place of any character, including blank	None
*	Asterisk, used to match a sub expression zero or multiple times before it	zo* can map to “z” and “zoo”.
+	Addition, used to match a sub expression one or multiple times before it	zo+ can map to “zo” and “zoo”, but not “z”.
-	Hyphen. It connects two values (the smaller one before it and the bigger one after it) to indicate a range together with [].	For example, “1-9” means numbers from 1 to 9 (inclusive); “a-h” means from a to h (inclusive).
[]	Selects one character from the group.	For example, [1-36A] can match only one character among 1, 2, 3, 6, and A.
()	A group of characters. It is usually used with “+” or “*”.	For example, (123A) means a string “123A”; “408(12)+” can match 40812 or 408121212. But it cannot match 408. That is, “12” can appear continuously and it must at least appear once.

Display functions

CLI offers the following feature:

When the information displayed exceeds one screen, you can pause using one of the methods shown in Table 85.

Table 85 Display functions

Action	Function
Press <Space> when information display pauses	Continues to display information of the next screen page.
Press <Enter> when information display pauses	Continues to display information of the next line.
Enter <Ctrl+C> when information display pauses	Stops the display and the command execution.
<Ctrl+E>	Moves the cursor to the end of the current line.
<PageUp>	Displays information on the previous page.
<PageDown>	Displays information on the next page.

Saving History Commands

The CLI can automatically save the commands that have been used. You can invoke and repeatedly execute them as needed. By default, the CLI can save up to ten commands for each user. You can use the **history-command max-size** command to set the capacity of the history commands log buffer for the current user interface (For the detailed description of the **history-command max-size** command, refer to "Controlling Login Users" on page 75). The following table lists the operations that you can perform.

Follow these steps to access history commands:

To do...	Use the key/command...	Result
View the history commands	display history-command	Displays the commands that you have entered
Access the previous history command	Up-arrow key or <Ctrl+P>	Displays the earlier history command, if there is any.
Access the next history command	Down-arrow key or <Ctrl+N>	Displays the next history command, if there is any.



You may use arrow keys to access history commands in Windows 200X and XP Terminal or Telnet. However, the up-arrow and down-arrow keys are invalid in Windows 9X HyperTerminal, because they are defined in a different way. You can use <Ctrl+P> and <Ctrl+N> instead.

Command Line Error Information

The commands are executed only if they have no syntax error. Otherwise, error information is reported. Table 86 lists some common errors.

Table 86 Common command line errors

Error information	Cause
% Unrecognized command found at '^' position.	The command was not found. The keyword was not found. Parameter type error The parameter value is beyond the allowed range.
% Incomplete command found at '^' position.	Incomplete command
% Ambiguous command found at '^' position.	Ambiguous command,
Too many parameters	Too many parameters
% Wrong parameter found at '^' position.	Wrong parameter

84

SYSTEM MAINTAINING AND DEBUGGING

When maintaining and debugging the system, go to these sections for information you are interested in:

- “System Maintaining and Debugging Overview” on page 1033
- “System Maintaining and Debugging” on page 1035
- “System Maintaining Example” on page 1036

System Maintaining and Debugging Overview

Introduction to System Maintaining and Debugging

You can use the **ping** command and the **tracert** command to verify the current network connectivity.

The ping command

You can use the **ping** command to verify whether a device with a specified address is reachable, and to examine network connectivity.

The **ping** command involves the following steps in its execution:

- 1 The source device sends an ICMP echo request to the destination device.
- 2 If the network is functioning properly, the destination device responds by sending an ICMP echo reply to the source device after receiving the ICMP echo request.
- 3 If there is network failure, the source device displays timeout or destination unreachable.
- 4 Display related statistics.

Output of the **ping** command includes:

- Information on the destination’s responses towards each ICMP echo request, if the source device has received the ICMP echo reply within the timeout time, it displays the number of bytes of the echo reply, the message sequence number, Time to Live (TTL), and the response time.
- If within the period set by the timeout timer, the destination device has not received the ICMP response, it displays the prompt information.
- The **ping** command can apply to the destination’s name or IP address. If the destination’s name is unknown, the prompt information is displayed.
- The statistics during the ping operation, which include number of packets sent, number of echo reply messages received, percentage of messages not received, the minimum, average, and maximum response time.

The **tracert** command

By using the **tracert** command, you can trace the routers involved in delivering a packet from source to destination. This is useful for identification of failed node(s) in the event of network failure.

The **tracert** command involves the following steps in its execution:

- 1 The source device sends a packet with a TTL value of 1 to the destination device.
- 2 The first hop (the router that first receives the packet) responds by sending a TTL-expired ICMP message to the source, with its IP address encapsulated. In this way, the source device can get the address of the first router.
- 3 The source device sends a packet with a TTL value of 2 to the destination device.
- 4 The second hop responds with a TTL-expired ICMP message, which gives the source device the address of the second router.
- 5 The above process continues until the ultimate destination device is reached. In this way, the source device can trace the addresses of all the routers that have been used to get to the destination device.

Introduction to System Debugging

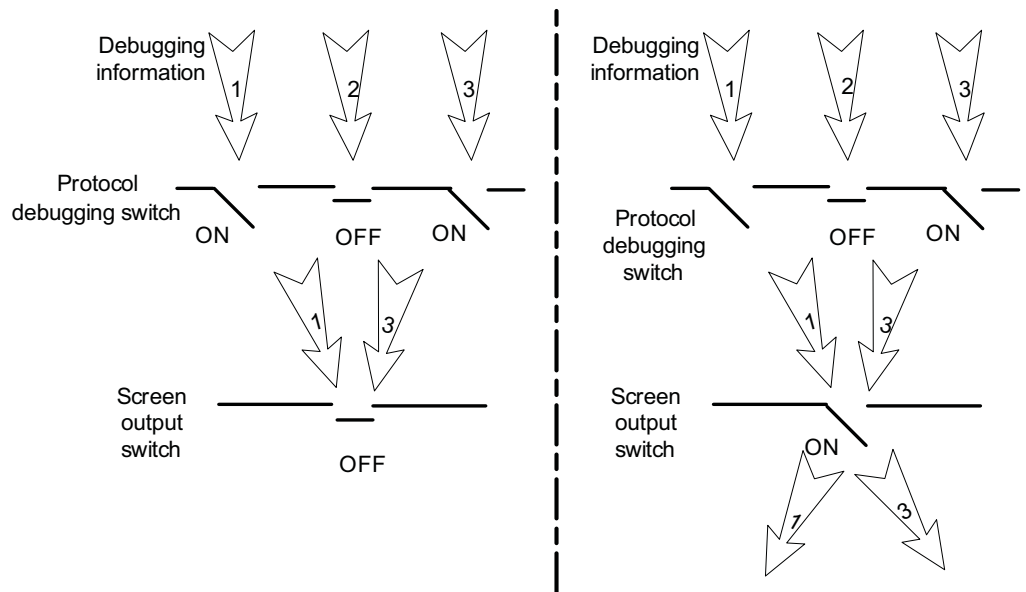
The device provides various debugging functions. For the majority of protocols and features supported, the system provides corresponding debugging information to help users diagnose errors.


The following two switches control the display of debugging information:

- Protocol debugging switch, which controls protocol-specific debugging information
- Screen output switch, which controls whether to display the debugging information on a certain screen.

Figure 305 illustrates the relationship between the protocol debugging switch and the screen output switch. Assume that the device can output debugging information to module 1, 2 and 3. Only when both are turned on can debugging information be output on a terminal.

Figure 305 The relationship between the protocol and screen debugging switch




 *Displaying debugging information on the terminal is the most commonly used way to output debugging information. You can also output debugging information to other directions. For details, refer to “Configuring Information Center” on page 1009.*

System Maintaining and Debugging

System Maintaining

To do...	Use the command...	Remarks
Check whether a specified IP address can be reached	ping [ip] [-a source-ip -c count -f -h ttl -i interface-type interface-number -m interval -n -p pad -q -r -s packet-size -t timeout -tos tos -v] * remote-system	Optional Used in IPv4 network Available in any view
	ping ipv6 [-a source-ipv6 -c count -m interval -s packet-size -t timeout] * remote-system [-i interface-type interface-number]	Optional Used in IPv6 network Available in any view
View the routes from the source to the destination	tracert [-a source-ip -f first-ttl -m max-ttl -p port -q packet-number -w timeout] * remote-system	Optional Used in IPv4 network Available in any view
	tracert ipv6 [-f first-ttl -m max-ttl -p port -q packet-number -w timeout] * remote-system	Optional Used in IPv6 network Available in any view

 ■ *For a low-speed network, you are recommended to set a larger value for the timeout timer (indicated by the -t parameter in the command) when configuring the **ping** command.*

- Only the directly connected segment address can be pinged if the outgoing interface is specified with the **-i** argument.

System Debugging

To do...	Use the command...	Remarks
Enable the terminal monitoring of system information	terminal monitor	Optional The terminal monitoring on the console is enabled by default and that on the monitoring terminal is disabled by default.
Enable the terminal display of debugging information	terminal debugging	Required Disabled by default Available in user view
Enable debugging for a specified module	debugging { all [timeout time] <i>module-name</i> [<i>option</i>] }	Required Disabled by default Available in user view
Display the enabled debugging functions	display debugging [interface <i>interface-type interface-number</i>] [<i>module-name</i>]	Optional Available in any view



- The **debugging** commands are usually used by administrators in diagnosing network failure.
- Output of the debugging information may reduce system efficiency, especially during execution of the **debugging all** command.
- After completing the debugging, you are recommended to use the **undo debugging all** command to disable all the debugging functions.
- You must configure the **debugging**, **terminal debugging** and **terminal monitor** commands first to display the detailed debugging information on the terminal. For the detailed description on the **terminal debugging** and **terminal monitor** commands, refer to “Configuring Information Center” on page 1009.

System Maintaining Example

Network requirements

- The IP address of the destination device is 10.1.1.4.
- Display the routers used while packets are forwarded from the current device to the destination device.

Network diagram (omitted here)

Configuration procedure

```
<Sysname> tracert 10.1.1.4
traceroute to 10.1.1.4 30 hops max, 40 bytes packet
 1 128.3.112.1 19 ms 19 ms 0 ms
 2 128.32.216.1 39 ms 39 ms 19 ms
 3 128.32.136.23 39 ms 40 ms 39 ms
 4 128.32.168.22 39 ms 39 ms 39 ms
 5 128.32.197.4 40 ms 59 ms 59 ms
 6 131.119.2.5 59 ms 59 ms 59 ms
```



```
7 129.140.70.13 99 ms 99 ms 80 ms
8 129.140.71.6 139 ms 239 ms 319 ms
9 129.140.81.7 220 ms 199 ms 199 ms
10 10.1.1.4 239 ms 239 ms 239 ms
```

The above output shows that nine routers are used from the source to the destination device.

When configuring device management, go to these sections for information you are interested in:

- "Device Management Overview" on page 1039
- "Configuring Device Management" on page 1039
- "Displaying and Maintaining Device Management Configuration" on page 1043
- "Device Management Configuration Example" on page 1043



File names in this document comply with the following rules

- *Path + file name (namely, a full file name): File on a specified path. A full file name consists of 1 to 135 characters.*
- *File name" (namely, only a file name without a path): File on the current working path. The file name without a path consists of 1 to 91 characters.*

Device Management Overview

Through the device management function, you can view the current working state of a device, configure running parameters, and perform daily device maintenance and management.

Currently, the following device management functions are available:

- "Rebooting a Device" on page 1039
- "Specifying a Boot ROM File for the Next Device Boot" on page 1040
- "Upgrading Boot ROM" on page 1040
- "Clearing the 16-bit Interface Indexes Not Used in the Current System" on page 1041
- "Identifying and Diagnosing Pluggable Transceivers" on page 1042

Configuring Device Management

Rebooting a Device

When a fault occurs to a running device, you can remove the fault by rebooting the device, depending on the actual situation. You can set a time at which the device can automatically reboot. You can also set a delay so that the device can automatically reboot in the delay.

Follow these steps to reboot a device:

To do...	Use the command...	Remarks
Reboot a device	reboot	Optional Available in user view.
Enable the scheduled reboot function and specify a specific reboot time and date	schedule reboot at <i>hh:mm</i> [<i>date</i>]	Optional The scheduled reboot function is disabled by default.
Enable the scheduled reboot function and specify a reboot waiting time	schedule reboot delay { <i>hh:mm</i> <i>mm</i> }	Execute the command in user view.

**CAUTION:**

- *The precision of the rebooting timer is 1 minute. One minute before the rebooting time, the device will prompt "REBOOT IN ONE MINUTE" and will reboot in one minute.*
- *The execution of the **reboot**, **schedule reboot at**, and **schedule reboot delay** commands can reboot a device. As a result, the ongoing services will be interrupted. Be careful to use these commands.*
- *If a primary boot file fails or does not exist, the device cannot be rebooted with this command. In this case, you can re-specify a primary boot file to reboot the device, or you can power off the device then power it on and the system automatically uses the secondary boot file to restart the device.*
- *Make sure that either the primary or the backup boot file or both are in normal use, when using the **schedule reboot** command to enable the scheduled reboot function.*
- *If you are performing file operations when the device is to be rebooted, the system does not execute the command for the sake of security.*

Specifying a Boot ROM File for the Next Device Boot

A Boot ROM file is an application file used to boot the device. When multiple Boot ROM files are available on the storage device, you can specify a file for the next device boot by executing the following command.

Follow these steps to specify a file for the next device boot:

To do...	Use the command...	Remarks
Specify a boot file for the device	boot-loader file <i>file-url</i> { main backup }	Required Available in user view,



CAUTION: *The file for the next device boot must be saved under the root directory of the device (for a device supporting storage device partition, the file must be saved on the first partition). You can copy or move a file to change the path of it to the root directory.*

Upgrading Boot ROM

During the operation of the device, you can use Boot ROM in the storage device to upgrade Boot ROM programs that are running on the device.

Since the Boot ROM programs vary with devices, users are easily confused and make serious mistakes when upgrading Boot ROM. After the validity check function is enabled, the device will strictly check the Boot ROM upgrade files for

correctness and version configuration information to ensure a successful upgrade. You are recommended to enable the validity check function before upgrading Boot ROM.

Follow these steps to upgrade Boot ROM:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the validity check function when upgrading Boot ROM	bootrom-update security-check enable	Optional Enabled by default.
Return to user view	quit	-
Upgrade the Boot ROM program of the device	bootrom update file file-url	Required Available in user view



Restart the device to validate the upgraded Boot ROM.

Clearing the 16-bit Interface Indexes Not Used in the Current System

In practical networks, the network management software requires the device to provide a uniform, stable 16-bit interface index. That is, a one-to-one relationship should be kept between the interface name and the interface index in the same device.

For the purpose of the stability of an interface index, the system will save the 16-bit interface index when a module or logical interface is removed.

If you repeatedly insert and remove different subcards or interface modules to create or delete a large amount of logical interface, the interface indexes will be used up, which will result in interface creation failures. To avoid such a case, you can clear all 16-bit interface indexes saved but not used in the current system in user view.

After the above operation,

- For a re-created interface, the new interface index may not be consistent with the original one.
- For existing interfaces, their interface indexes remain unchanged.

Follow the step below to clear the 16-bit interface indexes not used in the current system:

To do...	Use the command...	Remarks
Clear the 16-bit interface indexes saved but not used in the current system	reset unused porttag	Required Execute the command in user view.



CAUTION: A confirmation is required when you execute this command. If you fail to make a confirmation within 30 seconds or enter "N" to cancel the operation, the command will not be executed.

Identifying and Diagnosing Pluggable Transceivers

Introduction to pluggable transceivers

At present, four types of pluggable transceivers are commonly used, and they can be divided into optical transceivers and electrical transceivers based on transmission media as shown in Table 87.

Table 87 Commonly used pluggable transceivers

Transceiver type	Applied environment	Whether can be an optical transceiver	Whether can be an electrical transceiver
SFP (Small Form-factor Pluggable)	Generally used for 100M/1000M Ethernet interfaces or POS 155M/622M/2.5G interfaces	Yes	Yes
GBIC (GigaBit Interface Converter)	Generally used for 1000M Ethernet interfaces	Yes	Yes
XFP (10-Gigabit small Form-factor Pluggable)	Generally used for 10G Ethernet interfaces	Yes	No
XENPAK (10 Gigabit EtherNet Transceiver Package)	Generally used for 10G Ethernet interfaces	Yes	Yes



For pluggable transceivers supported by Switch 4800Gs-HI series Ethernet switches, refer to 3Com Switch 7750 Family, Switch 4800Gs-HI Series Installation Manuals.

Identifying pluggable transceivers

As pluggable transceivers are of various types and from different vendors, you can perform the following configurations to identify main parameters of the pluggable transceivers, including transceiver type, connector type, central wavelength of the laser sent, transfer distance and vendor name or vendor name specified.

Follow these steps to identify pluggable transceivers:

To do...	Use the command...	Remarks
Display main parameters of the pluggable transceiver(s)	display transceiver interface [<i>interface-type</i> <i>interface-number</i>]	Available for all pluggable transceivers
Display part of the electrical label information of the anti-spoofing transceiver(s) customized by 3Com	display transceiver manuinfo interface [<i>interface-type</i> <i>interface-number</i>]	Available for anti-spoofing pluggable transceiver(s) customized by 3Com only

- You can use the **Vendor Name** field in the prompt information of the **display transceiver interface** command to identify an anti-spoofing pluggable transceiver customized by 3Com. If the field is **3Com**, it is considered an 3Com-customized pluggable transceiver.
- Electrical label information is also called permanent configuration data or archive information, which is written to the storage device of a module during device debugging or test. The information includes name of the module, device serial number, and vendor name or vendor name specified.

Diagnosing pluggable transceivers

The system outputs alarm information for you to diagnose and troubleshoot faults of pluggable transceivers. Optical transceivers customized by 3Com also support the digital diagnosis function, which enables a transceiver to monitor the main

parameters such as temperature, voltage, laser bias current, TX power, and RX power. When these parameters are abnormal, you can take corresponding measures to prevent transceiver faults.

Follow these steps to display pluggable transceiver information:

To do...	Use the command...	Remarks
Display the current alarm information of the pluggable transceiver(s)	display transceiver alarm interface [<i>interface-type</i> <i>interface-number</i>]	Available for all pluggable transceivers
Display the currently measured value of the digital diagnosis parameters of the anti-spoofing optical transceiver(s) customized by 3Com	display transceiver diagnosis interface [<i>interface-type</i> <i>interface-number</i>]	Available for anti-spoofing pluggable optical transceiver(s) customized by 3Com only

Displaying and Maintaining Device Management Configuration

To do...	Use the command...	Remarks
Display the Boot ROM file used for the next boot	display boot-loader	Available in any view
Display the statistics of the CPU usage	display cpu-usage [<i>task number</i> [<i>offset</i>]] [verbose] [from-device]]	Available in any view
Display information about the device	display device [subslot <i>subslot-number</i> verbose]	Available in any view
Display manufacture information of the device	display device manuinfo	Available in any view
Display the temperature information of the device	display environment	Available in any view
Display the operating state of fans in a device	display fan [<i>fan-id</i>]	Available in any view
Display the usage of the memory of a device	display memory	Available in any view
Display the power state of a device	display power [<i>power-id</i>]	Available in any view
Display the reboot type of a device	display reboot-type	Available in any view
Display the reboot time of a device	display schedule reboot	Available in any view

Device Management Configuration Example

Remote Upgrade Configuration Example

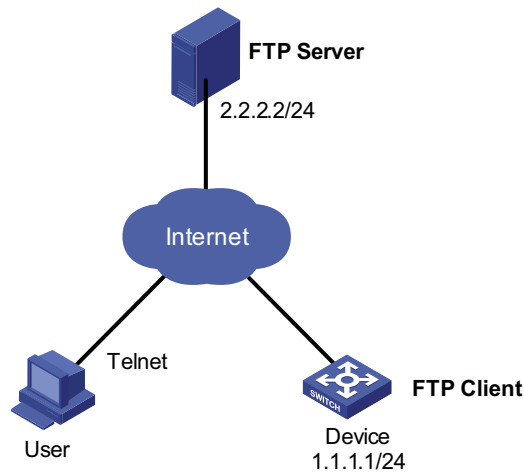
Network requirements

- Device serves as the FTP Client. The aaa.bin program and the boot.btm program are both saved under the aaa directory of the FTP Server.
- The IP address of a VLAN interface on Device is 1.1.1.1/24, the IP address of the FTP Server is 2.2.2.2/24, and FTP Server is reachable.
- User can log in to Device via Telnet to perform operations on Device (that is, download the application program from FTP Server and remotely upgrade

Device through command lines). Ensure that a route exists between User and Device.

Network diagram

Figure 306 Network diagram for remote upgrade



Configuration procedure

- Configuration on FTP Server (Note that configurations may vary with different types of servers)

Enable FTP Server.

```
<FTP-Server> system-view
[FTP-Server] ftp server enable
```

Set the FTP username to aaa and password to hello.

```
[FTP-Server] local-user aaa
[FTP-Server-luser-aaa] password cipher hello
```

Configure the user to have access to the aaa directory.

```
[FTP-Server-luser-aaa] service-type ftp ftp-directory flash:/aaa
```

- Configuration on Device



CAUTION: If the size of the Flash on the device is not large enough, delete the original application programs from the Flash before downloading.

Enter the following command in user view to log in to FTP Server.

```
<Device> ftp 2.2.2.2
Trying ...
Press CTRL+K to abort
Connected.
220 WFTPD 2.0 service (by Texas Imperial Software) ready for new use
r
User(none): aaa
331 Give me your password, please
Password:
```



```
230 Logged in successfully
[ftp]

# Download the aaa.bin and boot.btm programs on FTP Server to the Flash of
Device.

[ftp] get aaa.bin
[ftp] get boot.btm

# Clear the FTP connection and return to user view.

[ftp] bye
<Device>

# Enable the validity check function for Boot ROM file upgrade.

<Device> system-view
[Device] bootrom-update security-check enable
[Device] quit

# Upgrade the Boot ROM file of the device.

<Device> bootrom update file boot.btm

# Specify the application program for the next boot.

<Device> boot-loader file aaa.bin main

# Reboot the device. The application program is upgraded after the reboot.

<Device> reboot
Start to check configuration with next startup configuration file, please wait.....
This command will reboot the device. Current configuration will be lost in next startup if you continue. Continue? [Y/N]:y
This will reboot device. Continue? [Y/N]:y
```


When configuring NQA, go to these sections for information you are interested in:

- "NQA Overview" on page 1047
- "NQA Configuration Task List" on page 1050
- "Configuring the NQA Server" on page 1050
- "Enabling the NQA Client" on page 1051
- "Creating an NQA Test Group" on page 1051
- "Configuring an NQA Test Group" on page 1051
- "Configuring the Collaboration Function" on page 1061
- "Configuring Trap Delivery" on page 1061
- "Configuring Optional Parameters Common to an NQA Test Group" on page 1062
- "Scheduling an NQA Test Group" on page 1063
- "Displaying and Maintaining NQA" on page 1064
- "NQA Configuration Examples" on page 1064

NQA Overview

Introduction to NQA Network Quality Analyzer (NQA) analyzes network performance, services and service quality through sending test packets, and provides you with network performance and service quality parameters such as jitter, TCP connection delay, FTP connection delay and file transfer rate.

With the NQA test results, you can:

- 1 Know network performance in time and then take corresponding measures.
- 2 Diagnose and locate network faults.

Features of NQA **Supporting multiple test types**

Ping can use only the Internet Control Message Protocol (ICMP) to test the reachability of the destination host and the roundtrip time of a packet to the destination. NQA is an enhanced Ping tool used for testing the performance of protocols running on networks.

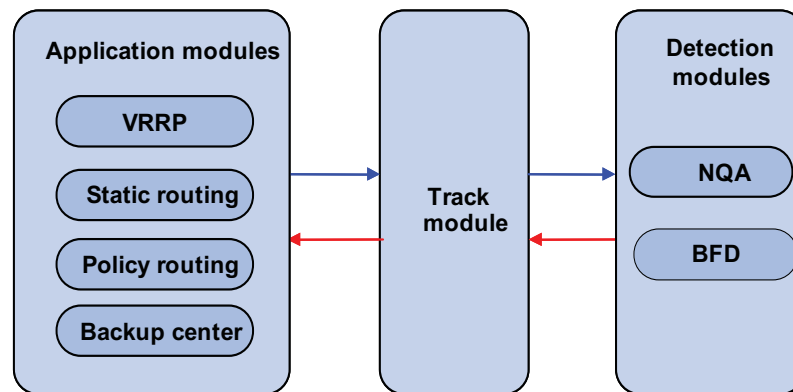
At present, NQA supports nine test types: ICMP-echo, DHCP, FTP, HTTP, UDP-jitter, SNMP, TCP, UDP-echo and DLSw.

In an NQA test, the client sends different types of test packets to the peer to detect the availability and the response time of the peer, helping you know protocol availability and network performance based on the test results.

Supporting the collaboration function

Collaboration is implemented by establishing collaboration entries to monitor the detection results of the current test group. If the number of consecutive probe failures reaches a certain limit, NQA's collaboration with other modules is triggered. The implementation of collaboration is shown in Figure 307.

Figure 307 Implementation of collaboration



The collaboration here involves three parts: the application modules, the Track module, and the detection modules.

- The detection modules monitor the link status, network performance and so on, and inform the Track module of the detection result.
- Upon receiving the detection result, the Track module changes the status of the Track object accordingly and informs the application modules. The Track module works between the application modules and the detection modules and is mainly used to obscure the difference of various detection modules to provide a unified interface for application modules.
- The application modules then deal with the changes accordingly based on the status of the Track object, and thus collaboration is implemented.

Take static routing as an example. You have configured a static route with the next hop 192.168.0.88. If 192.168.0.88 is reachable, the static route is valid; if 192.168.0.88 is unreachable, the static route is invalid. With the collaboration between NQA, Track module and application modules, real time monitoring of reachability of the static route can be implemented:

- 1 Monitor reachability of the destination 192.168.0.88 through NQA.
- 2 If 192.168.0.88 is detected to be unreachable, NQA will inform the static routing module through Track module.
- 3 The static routing module then can know that the static route is invalid.



For the detailed description of the Track module, refer to "Configuring Track-NQA Collaboration" on page 1238.

Supporting delivery of traps

Traps can be sent to the network management server when a test is completed, fails, or a probe fails.

A trap contains destination IP address, operation status, minimum and maximum Round Trip Time (RTT), probes sent, and time when the last probe is performed successfully. You can trace network running status with traps.

Basic Concepts of NQA Test group

NQA can test multiple protocols. A test group must be created for each type of NQA test and each test group can be related to only one type of NQA test.

Test and probe

After an NQA test is started, one test is performed at a regular interval and you can set the interval as needed.

One NQA test involves multiple consecutive probes and you can set the number of the probes.

In different test types, probe has different meanings:

- For a TCP or DLSw test, one probe means one connection;
- For a UDP-jitter test, the number of packets sent in one probe depends on the **probe packet-number** command;
- For an FTP, HTTP or DHCP test, one probe means to carry out a corresponding function;
- For an ICMP-echo or UDP-echo test, one packet is sent in one probe;
- For an SNMP test, three packets are sent in one probe.

NQA client and server

NQA client is the device initiating an NQA test and the NQA test group is created on the NQA client.

NQA server processes the test packets sent from the NQA client, as shown in Figure 308. The NQA server makes a response to the request originated by the NQA client by listening to the specified destination address and port number.

Figure 308 Relationship between the NQA client and NQA server



In most NQA tests, you only need to configure the NQA client; while in TCP, UDP-echo and UDP-jitter tests, you must configure the NQA server.

You can create multiple TCP or UDP listening services on the NQA server, with each listening service corresponding to a specified destination address and port number. The IP address and port number specified for a listening service on the

server must be consistent with those on the client and must be different from those of an existing listening service.

NQA Test Operation

After you create a test group and enter the test group view, you can configure related test parameters. Test parameters vary with the test type. For details, see the configuration procedure below.

To perform an NQA test successfully, make the following configurations on the NQA client:

- 1 Enable the NQA client;
- 2 Create a test group and configure test parameters according to the test type;
- 3 Perform the NQA test through the **nqa schedule** command.
- 4 View test results using the **display** or **debug** commands.

For TCP, UDP-jitter or UDP-echo tests, you need to configure the NQA server on the peer device.

NQA Configuration Task List

Complete these tasks to configure NQA.

Task	Remarks
"Configuring the NQA Server" on page 1050	Required for TCP, UDP-echo and UDP-jitter tests
"Enabling the NQA Client" on page 1051	Optional
"Creating an NQA Test Group" on page 1051	Required
"Configuring an NQA Test Group" on page 1051	Use any of the approaches.
"Configuring the ICMP-echo Test" on page 1051	
"Configuring the DHCP Test" on page 1053	
"Configuring the FTP Test" on page 1053	
"Configuring the HTTP Test" on page 1054	
"Configuring the UDP-jitter Test" on page 1055	
"Configuring the SNMP Test" on page 1057	
"Configuring the TCP Test" on page 1058	
"Configuring the UDP-echo Test" on page 1059	
"Configuring the DLSw Test" on page 1060	
"Configuring the Collaboration Function" on page 1061	Optional
"Configuring Trap Delivery" on page 1061	Optional
"Configuring Optional Parameters Common to an NQA Test Group" on page 1062	Optional
"Scheduling an NQA Test Group" on page 1063	Required

Configuring the NQA Server

Before performing TCP, UDP-echo or UDP-jitter tests, you need to configure the NQA server on the peer device. The NQA server makes a response to the request originated by the NQA client by listening to the specified destination address and port number.

Follow these steps to configure the NQA server:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the NQA server	nqa server enable	Required Disabled by default.
Configure the UDP or TCP listening function on the NQA server	nqa server { tcp-connect udp-echo } ip-address port-number	Required The IP address and port number must be consistent with those configured on the NQA client and must be different from those of an existing listening service.

Enabling the NQA Client

Configurations on the NQA client take effect only when the NQA client is enabled.

Follow these steps to enable the NQA client:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the NQA client	nqa agent enable	Optional Enabled by default.

Creating an NQA Test Group

One test corresponds to one test group. You can configure test types after you create a test group and enter the test group view.

Follow these steps to create an NQA test group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create an NQA test group and enter the NQA test group view	nqa entry admin-name operation-tag	Required



*If you execute the **nqa entry** command to enter the test group view with test type configured, you will enter the test type view of the test group directly.*

Configuring an NQA Test Group

Configuring the ICMP-echo Test

The ICMP test is used to test reachability of the destination host according to the ICMP-echo reply or timeout information.

Follow these steps to configure the ICMP-echo test:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Enter NQA test group view	nqa entry <i>admin-name</i> <i>operation-tag</i>	-
Configure the test type as ICMP-echo and enter test type view	type icmp-echo	Required
Configure the destination address for a test operation	destination ip <i>ip-address</i>	Required By default, no destination IP address is configured for a test operation.
Configure the size of probe packets sent	data-size <i>size</i>	Optional 100 bytes by default.
Configure the string used to fill a probe packet	data-fill <i>string</i>	Optional The string of fill characters of a probe packet is the string corresponding with the ASCII code 00 to 09 by default.
Specify the IP address of an interface as the source IP address of an ICMP-echo request	source interface <i>interface-type</i> <i>interface-number</i>	Optional By default, no interface address is specified as the source IP address of ICMP probe requests. If you use the source ip command to configure the source IP address of ICMP-echo probe requests, the source interface command is invalid. The interface specified by this command must be up. Otherwise, the probe will fail.
Configure the source IP address of a probe request	source ip <i>ip-address</i>	Optional By default, no source IP address is specified. If no source IP address is specified, but the source interface is specified, the IP address of the source interface is taken as the source IP address of ICMP probe requests. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the probe will fail.
Configure the next hop IP address for an ICMP-echo request	next-hop <i>ip-address</i>	Optional By default, no next hop IP address is configured.
Configure common optional parameters	Refer to "Configuring Optional Parameters Common to an NQA Test Group" on page 1062	Optional

Configuring the DHCP Test

The DHCP test is mainly used to test the existence of a DHCP server on the network as well as the time necessary for the DHCP server to respond to a client request and assign an IP address to the client.

Configuration prerequisites

Before performing a DHCP test, you need to configure the DHCP server. If the NQA (DHCP client) and the DHCP server are not in the same network segment, you need to configure a DHCP relay. For more information, refer to “DHCP Server Configuration” on page 797 and “DHCP Relay Agent Configuration” on page 813.

Configuring the DHCP test

Follow these steps to configure the DHCP test:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter NQA test group view	nqa entry <i>admin-name</i> <i>operation-tag</i>	-
Configure the test type as DHCP and enter test type view	type dhcp	Required
Specify an interface for a DHCP test	operation interface <i>interface-type interface-number</i>	Required By default, no interface is specified to perform a DHCP test. The interface specified by the source interface command must be up; otherwise, the test fails.
Configure common optional parameters	Refer to “Configuring Optional Parameters Common to an NQA Test Group” on page 1062	Optional



As DHCP test is a process to simulate address allocation in DHCP, the IP address of the interface performing the DHCP test will not be changed.

Configuring the FTP Test

The FTP test is mainly used to test the connection with a specified FTP server and the time necessary for the FTP client to transfer a file to or download a file from the FTP server.

Configuration prerequisites

Before the FTP test, you need to perform some configurations on the FTP server. For example, you need to configure the username and password used to log onto the FTP server. For the FTP server configuration, refer to “Configuring the FTP Server” on page 996.

Configuring the FTP test

Follow these steps to configure the FTP test:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Enter NQA test group view	nqa entry <i>admin-name</i> <i>operation-tag</i>	-
Configure the test type as FTP and enter test type view	type ftp	Required
Configure the destination address for a test operation	destination ip <i>ip-address</i>	Required By default, no destination IP address is configured for a test operation. The destination IP address for a test operation is the IP address of the FTP server.
Configure the source IP address of a probe request	source ip <i>ip-address</i>	Required By default, no source IP address is specified. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the test will fail.
Configure the operation type	operation { get put }	Optional By default, the operation type for the FTP is get , that is, obtaining files from the FTP server.
Configure a login username	username <i>name</i>	Required By default, no login username is configured.
Configure a login password	password <i>password</i>	Required By default, no login password is configured.
Specify a file to be transferred between the FTP server and the FTP client	filename <i>file-name</i>	Required By default, no file is specified.
Configure common optional parameters	Refer to "Configuring Optional Parameters Common to an NQA Test Group" on page 1062	Optional

Configuring the HTTP Test

The HTTP test is used to test the connection with a specified HTTP server and the time required to obtain data from the HTTP server.

Configuration prerequisites

Before performing an HTTP test, you need to configure the HTTP server.

Configuring the HTTP test

Follow these steps to configure the HTTP test:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter NQA test group view	nqa entry <i>admin-name</i> <i>operation-tag</i>	-

To do...	Use the command...	Remarks
Configure the test type as HTTP and enter test type view	type http	Required
Configure the destination address for a test operation	destination ip <i>ip-address</i>	Required By default, no destination IP address is configured for a test operation. The destination IP address for a test operation is the IP address of the HTTP server.
Configure the source IP address of a probe request	source ip <i>ip-address</i>	Optional By default, no source IP address is specified. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the test will fail.
Configure the operation type	operation { get post }	Optional By default, the operation type for the HTTP is get , that is, obtaining data from the HTTP server.
Configure the website that an HTTP test visits	url <i>url</i>	Required
Configure the HTTP version used in the HTTP test	http-version v1.0	Optional By default, HTTP 1.0 is used in an HTTP test.
Configure common optional parameters	Refer to "Configuring Optional Parameters Common to an NQA Test Group" on page 1062	Optional



The TCP port number for the HTTP server must be 80 in an HTTP test. Otherwise, the test will fail.

Configuring the UDP-jitter Test



You are not recommended to perform an NQA UDP-jitter test on ports from 1 to 1023 (known ports). Otherwise, the NQA test will fail or the corresponding services of this port will be unavailable.

Delay jitter refers to the difference between the interval of receiving two packets consecutively and the interval of sending these two packets. The procedure of a UDP-jitter test is as follows:

- The source sends packets at regular intervals to the destination port.
- The destination affixes a time stamp to each packet that it receives and then sends it back to the source.
- Upon receiving the packet, the source calculates the delay jitter, and the network status can be analyzed.

Configuration prerequisites

A UDP-jitter test requires cooperation between the NQA server and the NQA client. Before the UDP-jitter test, make sure that the UDP listening function is configured on the NQA server.

Configuring the UDP-jitter test

Follow these steps to configure the UDP-jitter test:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter NQA test group view	nqa entry <i>admin-name</i> <i>operation-tag</i>	-
Configure the test type as UDP-jitter and enter test type view	type udp-jitter	Required
Configure the destination address for a test operation	destination ip <i>ip-address</i>	Required By default, no destination IP address is configured for a test operation. The destination IP address must be consistent with that of the existing listening service on the NQA server.
Configure the destination port for a test operation	destination port <i>port-number</i>	Required By default, no destination port number is configured for a test operation. The destination port must be consistent with that of the existing listening service on the NQA server.
Specify the source port number for a request	source port <i>port-number</i>	Optional By default, no source port number is specified.
Configure the size of a probe packet sent	data-size <i>size</i>	Optional 100 bytes by default.
Configure the string of fill characters of a probe packet sent	data-fill <i>string</i>	Optional The string of fill characters of an ICMP probe packet is the string corresponding to the ASCII code 00 to 09 by default.
Configure the number of consecutive packets in a UDP-jitter probe	probe packet-number <i>packet-number</i>	Optional 10 by default.
Configure the interval for sending consecutive packets	probe packet-interval <i>packet-interval</i>	Optional 20 milliseconds by default.
Configure the time for waiting for a response in a UDP-jitter test	probe packet-timeout <i>packet-timeout</i>	Optional 3000 milliseconds by default.

To do...	Use the command...	Remarks
Configure the source IP address of a probe request in a test operation	source ip <i>ip-address</i>	Optional By default, no source IP address is specified. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the test will fail.
Configure common optional parameters	Refer to "Configuring Optional Parameters Common to an NQA Test Group" on page 1062	Optional



The number of probes made in a UDP-jitter test depends on the **probe count** command, while the number of probe packets sent in each probe depends on the **probe packet-number** command.

Configuring the SNMP Test

The SNMP query test is used to test the time the NQA client takes to send an SNMP query packet to the SNMP agent and then receive a response packet.

Configuration prerequisites

The SNMP agent function must be enabled on the device serving as an SNMP agent before the SNMP test. For the configuration of SNMP agent, refer to "SNMP Configuration" on page 933.

Configuring the SNMP test

Follow these steps to configure the SNMP test:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter NQA test group view	nqa entry <i>admin-name</i> <i>operation-tag</i>	-
Configure the test type as SNMP and enter test type view	type snmp	Required
Configure the destination address for a test operation	destination ip <i>ip-address</i>	Required By default, no destination IP address is configured for a test operation.
Specify the source port number for a probe request in a test operation	source port <i>port-number</i>	Optional By default, no source port number is specified.
Configure the source IP address of a probe request in a test operation	source ip <i>ip-address</i>	Optional By default, no source IP address is specified. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the test will fail.

To do...	Use the command...	Remarks
Configure common optional parameters	Refer to "Configuring Optional Parameters Common to an NQA Test Group" on page 1062	Optional

Configuring the TCP Test



You are not recommended to perform an NQA TCP test on ports from 1 to 1023 (known ports). Otherwise, the NQA test will fail or the corresponding services of this port will be unavailable.

The TCP test is used to test the TCP connection between the client and the specified server and the setup time for the connection.

Configuration prerequisites

A TCP test requires cooperation between the NQA server and the NQA client. The TCP listening function needs to be configured on the NQA server before the TCP test.

Configuring the TCP test

Follow these steps to configure the TCP test:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter NQA test group view	nqa entry <i>admin-name</i> <i>operation-tag</i>	-
Configure the test type as TCP and enter test type view	type tcp	Required
Configure the destination address for a test operation	destination ip <i>ip-address</i>	Required By default, no destination IP address is configured for a test operation. The destination address must be the IP address of the listening service configured on the NQA server.
Configure the destination port	destination port <i>port-number</i>	Required By default, no destination port number is configured for a test operation. The destination port number must be consistent with port number of the listening service configured on the NQA server.

To do...	Use the command...	Remarks
Configure the source IP address of a probe request in a test operation	source ip <i>ip-address</i>	Optional By default, no source IP address is specified. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the test will fail.
Configure common optional parameters	Refer to "Configuring Optional Parameters Common to an NQA Test Group" on page 1062	Optional

Configuring the UDP-echo Test



You are not recommended to perform an NQA UDP test on ports from 1 to 1023 (known ports). Otherwise, the NQA test will fail or the corresponding services of this port will be unavailable.

The UDP-echo test is used to test the roundtrip time of a UDP-echo packet from the client to the specified server.

Configuration prerequisites

A UDP-echo test requires cooperation between the NQA server and the NQA client. The UDP listening function needs to be configured on the NQA server before the UDP-echo test.

Configuring the UDP-echo test

Follow these steps to configure the UDP-echo test

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter NQA test group view	nqa entry <i>admin-name operation-tag</i>	-
Configure the test type as UDP-echo and enter test type view	type udp-echo	Required
Configure the destination address for a test operation	destination ip <i>ip-address</i>	Required By default, no destination IP address is configured for a test operation. The destination address must be the IP address of the listening service configured on the NQA server.

To do...	Use the command...	Remarks
Configure the destination port	destination port <i>port-number</i>	Required By default, no destination port number is configured for a test operation. The destination port number must be the port number of the listening service configured on the NQA server.
Configure the size of probe packets sent	data-size <i>size</i>	Optional 100 bytes by default.
Configure the string of fill characters of a probe packet sent	data-fill <i>string</i>	Optional The string of fill characters of an ICMP probe packet is the string corresponding with the ASCII code 00 to 09 by default.
Specify a source port number for a probe request in a test operation	source port <i>port-number</i>	Optional By default, no source port number is specified.
Configure the source IP address of a probe request in a test operation	source ip <i>ip-address</i>	Optional By default, no source IP address is specified. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the test will fail.
Configure common optional parameters	Refer to "Configuring Optional Parameters Common to an NQA Test Group" on page 1062	Optional

Configuring the DLSw Test

The DLSw test is used to test the response time of the DLSw device.

Configuration prerequisites

Enable the DLSw function on the peer device before DLSw test.

Configuring the DLSw test

Follow these steps to configure the DLSw test:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter NQA test group view	nqa entry <i>admin-name</i> <i>operation-tag</i>	-
Configure the test type as DLSw and enter test type view	type dlsw	Required
Configure the destination address for a test operation	destination ip <i>ip-address</i>	Required By default, no destination IP address is configured for a test operation.

To do...	Use the command...	Remarks
Configure the source IP address of a probe request in a test operation	source ip <i>ip-address</i>	Optional By default, no source IP address is specified. The source IP address must be that of an interface on the device and the interface must be up. Otherwise, the test will fail.
Configure common optional parameters	Refer to "Configuring Optional Parameters Common to an NQA Test Group" on page 1062	Optional

Configuring the Collaboration Function

Collaboration is implemented by establishing collaboration entries to monitor the detection results of the current test group. If the number of consecutive probe failures reaches the threshold, the configured action is triggered.

Follow these steps to configure the collaboration function:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter NQA test group view	nqa entry <i>admin-name</i> <i>operation-tag</i>	-
Enter test type view of the test group	type { dhcp dls w ftp http icmp-echo snmp tcp udp-echo }	The collaboration function is not supported in UDP-jitter tests.
Create a Reaction entry	reaction <i>item-num</i> checked-element probe-fail threshold-type consecutive <i>occurrences</i> [action-type { none trigger-only }]	Required Not created by default.
Exit to system view	quit	-
Create a Track object and associate it with the specified Reaction entry of the NQA test group	track <i>entry-number</i> nqa entry <i>admin-name</i> <i>operation-tag</i> reaction <i>item-num</i>	Required Not created by default.



CAUTION: You cannot modify the content of a reaction entry using the **reaction** command after the collaboration object is created.

Configuring Trap Delivery

Traps can be sent to the network management server when test is completed, test fails or probe fails.

Configuration prerequisites

Before configuring trap delivery, you need to configure the destination address of the trap message with the **snmp-agent target-host** command, create an NQA test group, and configure related parameters. For the introduction to the **snmp-agent target-host** command, refer to "SNMP Configuration" on page 931.

Configuring trap delivery

Follow these steps to configure trap delivery:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter NQA test group view	nqa entry <i>admin-name</i> <i>operation-tag</i>	-
Enter test type view of the test group	type { dhcp dls w ftp http icmp-echo snmp tcp udp-echo udp-jitter }	-
Configure to send traps to network management server under specified conditions	reaction trap { probe-failure <i>consecutive-probe-failures</i> test-complete test-failure <i>cumulate-probe-failures</i> }	Optional No traps are sent to the network management server by default.

Configuring Optional Parameters Common to an NQA Test Group

Optional parameters common to an NQA test group are valid only for tests in this test group.

Unless otherwise specified, the following parameters are applicable to all test types and they can be configured according to the actual conditions.

Follow these steps to configure optional parameters common to an NQA test group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter NQA test group view	nqa entry <i>admin-name</i> <i>operation-tag</i>	-
Enter test type view of a test group	type { dhcp dls w ftp http icmp-echo snmp tcp udp-echo udp-jitter }	-
Configure the descriptive string for a test group	description <i>string</i>	Optional By default, no descriptive string is available for a test group.
Configure the interval between two consecutive tests for a test group	frequency <i>interval</i>	Optional By default, the interval between two consecutive tests for a test group is 0 milliseconds, that is, only one test is performed. If the last test is not completed when the interval specified by the frequency command is reached, a new test is not started.
Configure the number of probes in a test	probe count <i>times</i>	Optional By default, one probe is performed in a test.

To do...	Use the command...	Remarks
Configure the NQA probe timeout time	probe timeout <i>timeout</i>	Optional By default, the timeout time is 3000 milliseconds. This parameter is not available for a UDP-jitter test.
Configure the maximum number of history records that can be saved in a test group	history-records <i>number</i>	Optional 50 by default.
Configure the maximum number of hops a probe packet traverses in the network	ttl <i>value</i>	Optional 20 by default. This parameter is not available for a DHCP test.
Configure the ToS field in an IP packet header in an NQA probe packet	tos <i>value</i>	Optional 0 by default. This parameter is not available for a DHCP test.
Enable the routing table bypass function	route-option bypass-route	Optional Disabled by default. This parameter is not available for a DHCP test.

Scheduling an NQA Test Group

With this configuration, you can set the start time and time period for a test group to perform the test and start the test.

Configuration prerequisites

Before scheduling an NQA test group, make sure:

- Required test parameters corresponding to a test type have been configured;
- For the test which needs the cooperation with the NQA server, configuration on the NQA server has been completed.

Scheduling an NQA test group

Follow these steps to schedule an NQA test group:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Schedule an NQA test group	nqa schedule <i>admin-name operation-tag start-time now lifetime forever</i>	Required
Configure the maximum number of the tests that the NQA client can simultaneously perform	nqa agent max-concurrent <i>number</i>	Optional The default number is 2



CAUTION: After an NQA test group is scheduled, you cannot enter the test group view or test type view.

Displaying and Maintaining NQA

To do...	Use the command...
Display NQA test operation information	display nqa { result history } [<i>admin-name</i> <i>operation-tag</i>]
Display NQA server status	display nqa server status

NQA Configuration Examples

ICMP-echo Test Configuration Example

Network requirements

Use the NQA ICMP function to test whether the NQA client (Device A) can send packets to the specified destination (Device B) and test the roundtrip time of packets.

Network diagram

Figure 309 Network diagram for the ICMP-echo test



Configuration procedure

Create an ICMP-echo test group and configure related test parameters.

```

<DeviceA> system-view
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type icmp-echo
[DeviceA-nqa-admin-test-icmp-echo] destination ip 10.2.2.2
  
```

Configure optional parameters.

```

[DeviceA-nqa-admin-test-icmp-echo] probe count 10
[DeviceA-nqa-admin-test-icmp-echo] probe timeout 500
[DeviceA-nqa-admin-test-icmp-echo] quit
  
```

Enable the ICMP-echo test operation.

```

[DeviceA] nqa schedule admin test start-time now lifetime forever
  
```

Display results of an ICMP-echo test.

```

[DeviceA] display nqa result admin test
NQA entry(admin admin, tag test) test results:
  Destination IP address: 10.2.2.2
  Send operation times: 10          Receive response times: 10
  Min/Max/Average round trip time: 0/16/1
  Square-Sum of round trip time: 256
  Last succeeded probe time: 2007-03-14 17:21:07.8
Extend results:
  Packet lost in test: 0%
  Failures due to timeout: 0
  Failures due to disconnect: 0
  Failures due to no connection: 0
  
```

```
Failures due to sequence error: 0
Failures due to internal error: 0
Failures due to other errors: 0
```

DHCP Test Configuration Example

Network requirements

Use the NQA DHCP function to test the time necessary for Switch A to obtain an IP address from the DHCP server Switch B.

Network diagram

Figure 310 Network diagram for DHCP



Configuration procedure

Create a DHCP test group and configure related test parameters.

```
<SwitchA> system-view
[SwitchA] nqa entry admin test
[SwitchA-nqa-admin-test] type dhcp
[SwitchA-nqa-admin-test-dhcp] operation interface vlan-interface 2
[SwitchA-nqa-admin-test-dhcp] quit
```

Enable the DHCP test.

```
[SwitchA] nqa schedule admin test start-time now lifetime forever
```

Display results of one DHCP test.

```
[SwitchA] display nqa result admin test
NQA entry(admin admin, tag test) test results:
  Send operation times: 1          Receive response times: 1
  Min/Max/Average round trip time: 624/624/624
  Square-Sum of round trip time: 389376
  Last succeeded probe time: 2007-03-14 17:47:29.3
Extend results:
  Packet lost in test: 0%
  Failures due to timeout: 0
  Failures due to disconnect: 0
  Failures due to no connection: 0
  Failures due to sequence error: 0
  Failures due to internal error: 0
  Failures due to other errors: 0
```

FTP Test Configuration Example

Network requirements

Use the NQA FTP function to test the connection with a specified FTP server and the time necessary for Device A to upload a file to the FTP server. The login username is **admin**, the login password is **systemtest**, and the file to be transferred to the FTP server is **config.txt**.

Network diagram

Figure 311 Network diagram for FTP



Configuration procedure

Create an FTP test group and configure related test parameters.

```
<DeviceA> system-view
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type ftp
[DeviceA-nqa-admin-test-ftp] destination ip 10.2.2.2
[DeviceA-nqa-admin-test-ftp] source ip 10.1.1.1
[DeviceA-nqa-admin-test-ftp] operation put
[DeviceA-nqa-admin-test-ftp] username admin
[DeviceA-nqa-admin-test-ftp] password systemtest
[DeviceA-nqa-admin-test-ftp] filename config.txt
[DeviceA-nqa-admin-test-ftp] quit
```

Enable the FTP test.

```
[DeviceA] nqa schedule admin test start-time now lifetime forever
```

Display results of an FTP test.

```
[DeviceA] display nqa result admin test
NQA entry(admin admin, tag test) test results:
  Destination IP address: 10.2.2.2
  Send operation times: 1                Receive response times: 1
  Min/Max/Average round trip time: 173/173/173
  Square-Sum of round trip time: 29929
  Last succeeded probe time: 2007-03-14 13:28:48.5
Extend results:
  Packet lost in test: 0%
  Failures due to timeout: 0
  Failures due to disconnect: 0
  Failures due to no connection: 0
  Failures due to sequence error: 0
  Failures due to internal error: 0
  Failures due to other errors: 0
```

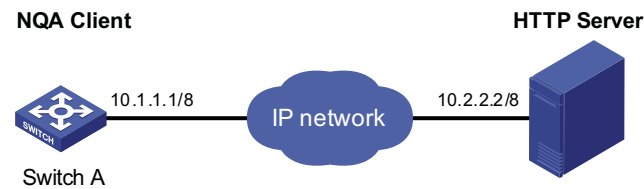
HTTP Test Configuration Example

Network requirements

Use the HTTP function to test the connection with a specified HTTP server and the time required to obtain data from the HTTP server.

Network diagram

Figure 312 Network diagram for the HTTP test



Configuration procedure

Create an HTTP test group and configure related test parameters.

```

<DeviceA> system-view
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type http
[DeviceA-nqa-admin-test-http] destination ip 10.2.2.2
[DeviceA-nqa-admin-test-http] operation get
[DeviceA-nqa-admin-test-http] url /index.htm
[DeviceA-nqa-admin-test-http] http-version v1.0
[DeviceA-nqa-admin-test-http] quit
  
```

Enable the HTTP test.

```
[DeviceA] nqa schedule admin test start-time now lifetime forever
```

Display results of an HTTP test.

```

[DeviceA] display nqa result admin test
NQA entry(admin admin, tag test) test results:
  Destination IP address: 10.2.2.2
    Send operation times: 1                Receive response times: 1
    Min/Max/Average round trip time: 64/64/64
    Square-Sum of round trip time: 4096
    Last succeeded probe time: 2007-03-27 13:40:36.2
Extend results:
  Packet lost in test: 0%
  Failures due to timeout: 0
  Failures due to disconnect: 0
  Failures due to no connection: 0
  Failures due to sequence error: 0
  Failures due to internal error: 0
  Failures due to other errors: 0
  
```

UDP-jitter Test Configuration Example

Network requirements

Use the NQA UDP-jitter function to test the delay jitter of packet transmission between Device A and Device B.

Network diagram

Figure 313 Network diagram for UDP-jitter test



Configuration procedure

1 Configure Device B.

Enable the NQA server and configure the listening IP address as 10.2.2.2 and port number as 9000.

```

<DeviceB> system-view
[DeviceB] nqa server enable
[DeviceB] nqa server udp-echo 10.2.2.2 9000
  
```

1 Configure Device A.

Create a UDP-jitter test group and configure related test parameters.

```

<DeviceA> system-view
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type udp-jitter
[DeviceA-nqa-admin-test-udp-jitter] destination ip 10.2.2.2
[DeviceA-nqa-admin-test-udp-jitter] destination port 9000
[DeviceA-nqa-admin-test-udp-jitter] quit
  
```

Enable the UDP-jitter test.

```

[DeviceA] nqa schedule admin test start-time now lifetime forever
  
```

Display results of a UDP-jitter test.

```

[DeviceA] display nqa result admin test
NQA entry(admin admin, tag test) test results:
  Destination IP address: 10.2.2.2
  Send operation times: 10          Receive response times: 10
  Min/Max/Average round trip time: 31/47/32
  Square-Sum of round trip time: 10984
  Last succeeded probe time: 2007-04-29 20:05:49.1
Extend results:
  Packet lost in test: 0%
  Failures due to timeout: 0
  Failures due to disconnect: 0
  Failures due to no connection: 0
  Failures due to sequence error: 0
  Failures due to internal error: 0
  Failures due to other errors: 0
UDP-jitter results:
  RTT number: 10
  SD max delay: 23
  Min positive SD: 1
  Max positive SD: 1
  Positive SD number: 2
  Positive SD sum: 2
  Positive SD average: 1
  Positive SD square sum: 2
  DS max delay: 23
  Min positive DS: 1
  Max positive DS: 1
  Positive DS number: 2
  Positive DS sum: 16
  Positive DS average: 8
  Positive DS square sum: 226
  
```



```

Min negative SD: 1
Max negative SD: 15
Negative SD number: 3
Negative SD sum: 17
Negative SD average: 6
Negative SD square sum: 227
SD lost packet(s): 0
Lost packet(s) for unknown reason: 0

Min negative DS: 1
Max negative DS: 1
Negative DS number: 3
Negative DS sum: 17
Negative DS average: 6
Negative DS square sum: 227
DS lost packet(s): 0

```

SNMP Test Configuration Example

Network requirements

Use the NQA SNMP query function to test the time it takes Device A to send an SNMP query packet to the SNMP agent and receive a response packet.

Network diagram

Figure 314 Network diagram for SNMP test



Configuration procedure

1 Configurations on SNMP agent.

Enable the SNMP agent service and set the SNMP version to **all**, the read community to **public**, and the write community to **private**.

```

<DeviceB> system-view
[DeviceB] snmp-agent sys-info version all
[DeviceB] snmp-agent community read public
[DeviceB] snmp-agent community write private

```

2 Configurations on Device A.

Create an SNMP query test group and configure related test parameters.

```

<DeviceA> system-view
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type snmp
[DeviceA-nqa-admin-test-snmp] destination ip 10.2.2.2
[DeviceA-nqa-admin-test-snmp] quit

```

Enable the SNMP query test.

```

[DeviceA] nqa schedule admin test start-time now lifetime forever

```

Display results of an SNMP test.

```

[DeviceA] display nqa result admin test
NQA entry(admin admin, tag test) test results:
  Destination IP address: 10.2.2.2
  Send operation times: 1                      Receive response times: 1
  Min/Max/Average round trip time: 50/50/50
  Square-Sum of round trip time: 2500
  Last succeeded probe time: 2007-03-27 13:59:43.1
Extend results:

```

```

Packet lost in test: 0%
Failures due to timeout: 0
Failures due to disconnect: 0
Failures due to no connection: 0
Failures due to sequence error: 0
Failures due to internal error: 0
Failures due to other errors: 0

```

TCP Test Configuration Example

Network requirements

Use the NQA TCP function to test the time for establishing a TCP connection between Device A and Device B. The port number used is 9000.

Network diagram

Figure 315 Network diagram for the TCP test



Configuration procedure

1 Configure Device B.

Enable the NQA server and configure the listening IP address as 10.2.2.2 and port number as 9000.

```

<DeviceB> system-view
[DeviceB] nqa server enable
[DeviceB] nqa server tcp-connect 10.2.2.2 9000

```

2 Configure Device A.

Create a TCP test group and configure related test parameters.

```

<DeviceA> system-view
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type tcp
[DeviceA-nqa-admin-test-tcp] destination ip 10.2.2.2
[DeviceA-nqa-admin-test-tcp] destination port 9000
[DeviceA-nqa-admin-test-tcp] quit

```

Enable the TCP test.

```

[DeviceA] nqa schedule admin test start-time now lifetime forever

```

Display results of one TCP test.

```

[DeviceA] display nqa result admin test
NQA entry(admin admin, tag test) test results:
  Destination IP address: 10.2.2.2
  Send operation times: 1          Receive response times: 1
  Min/Max/Average round trip time: 13/13/13
  Square-Sum of round trip time: 169
  Last succeeded probe time: 2000-04-27 14:03:20.1

```

```

Extend results:
  Packet lost in test: 0%
  Failures due to timeout: 0
  Failures due to disconnect: 0
  Failures due to no connection: 0
  Failures due to sequence error: 0
  Failures due to internal error: 0
  Failures due to other errors: 0

```

UDP-echo Test Configuration Example

Network requirements

Use the NQA UDP-echo function to test the round trip time between Device A and Device B. The port number is 8000.

Network diagram

Figure 316 Network diagram for the UDP-echo test



Configuration procedure

1 Configure Device B.

Enable the NQA server and configure the listening IP address as 10.2.2.2 and port number as 8000.

```

<DeviceB> system-view
[DeviceB] nqa server enable
[DeviceB] nqa server udp-echo 10.2.2.2 8000

```

2 Configure Device A.

Create a UDP-echo test group and configure related test parameters.

```

<DeviceA> system-view
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type udp-echo
[DeviceA-nqa-admin-test-udp-echo] destination ip 10.2.2.2
[DeviceA-nqa-admin-test-udp-echo] destination port 8000
[DeviceA-nqa-admin-test-udp-echo] quit

```

Enable the UDP-echo test.

```

[DeviceA] nqa schedule admin test start-time now lifetime forever

```

Display results of one UDP-echo test.

```

[DeviceA] display nqa result admin test
NQA entry(admin admin, tag test) test results:
  Destination IP address: 10.2.2.2
  Send operation times: 1                      Receive response times: 1
  Min/Max/Average round trip time: 25/25/25
  Square-Sum of round trip time: 625
  Last succeeded probe time: 2007-03-27 14:07:40.7

```

```

Extend results:
  Packet lost in test: 0%
  Failures due to timeout: 0
  Failures due to disconnect: 0
  Failures due to no connection: 0
  Failures due to sequence error: 0
  Failures due to internal error: 0
  Failures due to other errors: 0

```

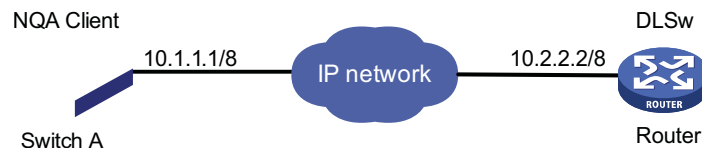
DLSw Test Configuration Example

Network requirements

Use the NQA DLSw function to test the response time of the DLSw device.

Network diagram

Figure 317 Network diagram for the DLSw test



Configuration procedure

Create a DLSw test group and configure related test parameters.

```

<DeviceA> system-view
[DeviceA] nqa entry admin test
[DeviceA-nqa-admin-test] type dlsw
[DeviceA-nqa-admin-test-dlsw] destination ip 10.2.2.2
[DeviceA-nqa-admin-test-dlsw] quit

```

Enable the DLSw test.

```

[DeviceA] nqa schedule admin test start-time now lifetime forever

```

Display results of one DLSw test.

```

[DeviceA] display nqa result admin test
NQA entry(admin admin, tag test) test results:
  Destination IP address: 10.2.2.2
  Send operation times: 1          Receive response times: 1
  Min/Max/Average round trip time: 19/19/19
  Square-Sum of round trip time: 361
  Last succeeded probe time: 2007-03-27 15:32:48.5
Extend results:
  Packet lost in test: 0%
  Failures due to timeout: 0
  Failures due to disconnect: 0
  Failures due to no connection: 0
  Failures due to sequence error: 0
  Failures due to internal error: 0
  Failures due to other errors: 0

```

87

VRRP CONFIGURATION

When configuring VRRP, go to these sections for information you are interested in:

- "Introduction to VRRP" on page 1073
- "Configuring VRRP for IPv4" on page 1081
- "Configuring VRRP for IPv6" on page 1084
- "IPv4-Based VRRP Configuration Examples" on page 1088
- "IPv6-Based VRRP Configuration Examples" on page 1096
- "Troubleshooting VRRP" on page 1105



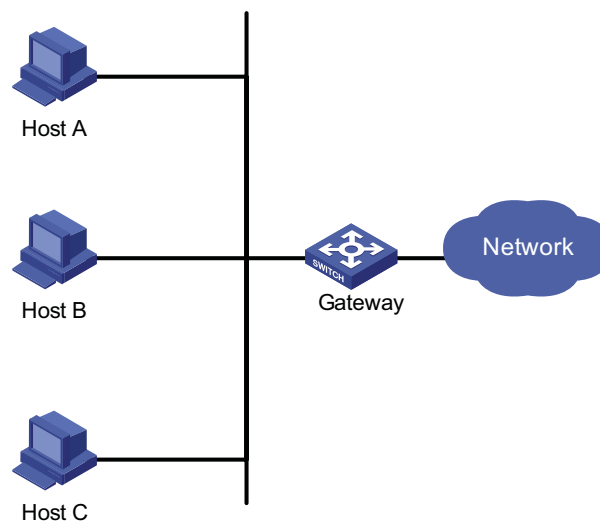
At present, the interfaces that VRRP involves can only be VLAN interfaces unless otherwise specified.

Introduction to VRRP

VRRP Overview

Normally, as shown in Figure 318, you can configure a default route with the gateway as the next hop for every host on a network segment, allowing all packets destined to the other network segments to be sent over the default route to the gateway and then be forwarded by the gateway. This enables hosts on a network segment to communicate with external networks. However, when the gateway fails, all the hosts using the gateway as the default next-hop switch fail to communicate with the external network.

Figure 318 LAN networking



Apparently, this approach to enabling hosts on a network to communicate with external networks is easy to configure but it imposes a very high requirement of performance stability on the device acting as the gateway. A common way to improve system reliability is to use more egress gateways, introducing the problem of routing among the multiple egresses.

Virtual Router Redundancy Protocol (VRRP) is an error-tolerant protocol designed to address this problem through separating physical devices from logical devices. Deploying VRRP on multicast and broadcast LANs such as Ethernet, you can ensure that the system can still provide highly reliable default links without changing configurations (such as dynamic routing protocols, route discovery protocols) when a device fails and prevent network interruption due to a single link failure.

There are two VRRP versions: VRRPv2 and VRRPv3. VRRPv2 is based on IPv4, while VRRPv3 is based on IPv6. The two versions implement the same functions but provide different commands.

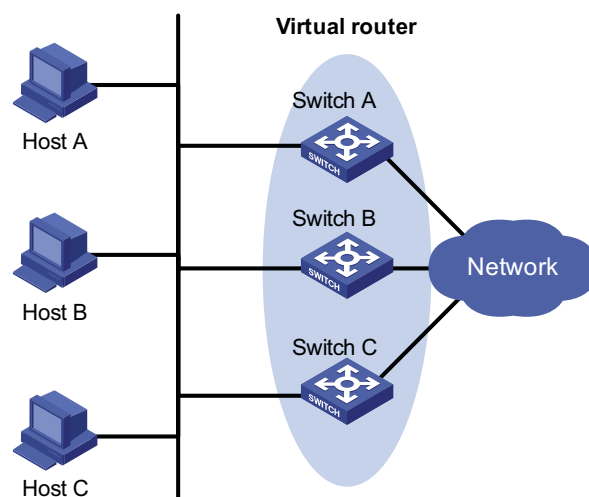
VRRP Standby Group Overview

VRRP combines a group of switches (including a master and multiple backups) on a LAN into a virtual router called standby group.

The VRRP standby group has the following features:

- A virtual router has an IP address. A host on the LAN only needs to know the IP address of the virtual router and uses the IP address as the next hop of the default route.
- Every host on the LAN communicates with external networks through the virtual router.
- Switches in the standby group elect the gateway according to their priorities. Once the master switch acting as the gateway fails, the other switches in the standby group elect a new gateway to undertake the responsibility of the failed switch, thus ensuring that the hosts in the network segment can communicate with the external networks uninterruptedly.

Figure 319 Network diagram for VRRP



As shown in Figure 319, Switch A, Switch B, and Switch C form a virtual router, which has its own IP address. Hosts on the Ethernet use the virtual router as the default gateway.

The switch with the highest priority of the three switches is elected as the master switch to act as the gateway, and the other two are backup switches.



CAUTION:

- *The IP address of the virtual router can be either an unused IP address on the segment where*
- *the standby group resides or the IP address of an interface on a switch in the standby group. In the latter case, the switch is called the IP address owner.*
- *In a VRRP standby group, there can only be one IP address owner.*

VRRP priority

VRRP determines the role (master or backup) of each switch in the standby group by priority. A switch with a higher priority has more opportunity to become the master.

VRRP priority is in the range of 0 to 255. A bigger number means a higher priority. Priorities 1 to 254 are configurable. Priority 0 is reserved for special uses and priority 255 for the IP address owner. When a switch acts as the IP address owner, its priority remains 255. That is, if there is an IP address owner in a standby group, it acts as the master as long as it works properly.

Working mode

A switch in a standby group can work in one of the following two modes:

- Non-preemption mode

Once a switch in the standby group becomes the master, it stays as the master as long as it operates normally, even if a backup switch is assigned a higher priority later.

- Preemption mode

Once a backup switch finds its priority higher than that of the switch acting as the master, it sends VRRP advertisements to start a new master switch election in the standby group and becomes the master. Accordingly, the original master switch becomes a backup.

Authentication mode

VRRP provides two authentication modes:

- **simple:** Simple text authentication

You can adopt the simple text authentication mode in a network facing possible security problems. A switch sending a packet fills the authentication key into the packet, and the switch receiving the packet compares its local authentication key with that of the received packet. If the two authentication keys are the same, the received VRRP packet is considered real and valid; otherwise, the received packet is considered an invalid one.

- **md5**: MD5 authentication

You can adopt MD5 authentication in a network facing severe security problems. The switch encrypts a packet to be sent using the authentication key and MD5 algorithm and saves the encrypted packet in the authentication header. The switch receiving the packet uses the authentication key to decrypt the packet and checks whether the packet is valid.

On a secure network, you need not set the authentication mode.

VRRP Timers VRRP timers include VRRP advertisement interval timer and VRRP preemption delay timer.

VRRP advertisement interval timer

The master switch in a VRRP standby group sends VRRP advertisements periodically to inform the other switches in the standby group that it operates properly.

You can adjust the interval of sending VRRP advertisements by setting the VRRP advertisement interval timer. If a backup switch receives no advertisements in three times the interval, the backup switch regards itself as the master switch and sends VRRP advertisements to start a new master switch election.

VRRP preemption delay timer

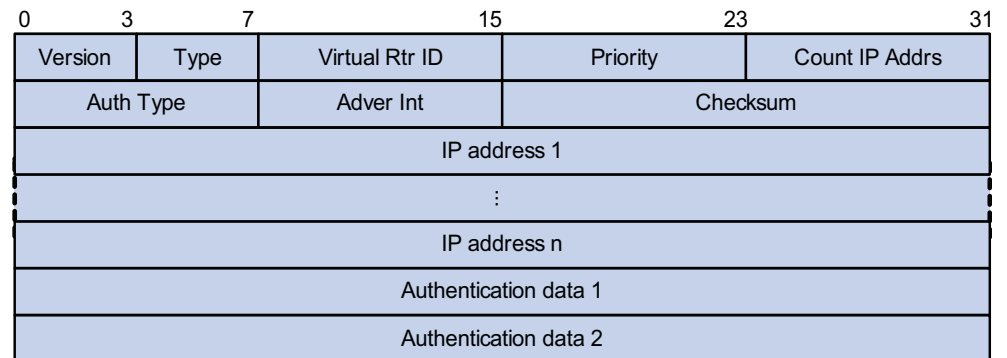
In an unstable network, a backup switch may fail to receive the packets from the master switch due to network congestion, thus causing the members in the group to change their states frequently. This problem can be addressed through setting the VRRP preemption delay timer.

With the VRRP preemption delay timer set, if a backup switch receives no advertisement in three times the advertisement interval and then in preemption delay, it considers that the master fails. In this case, it regards itself as the master and sends VRRP advertisements to start a new master switch election in a standby group.

Format of VRRP Packets VRRP uses multicast packets. The switch acting as the master sends VRRP packets periodically to declare its existence. VRRP packets are also used for checking the parameters of the virtual router and electing the master.

IPv4-based VRRP packet format

Figure 320 IPv4-based VRRP packet format

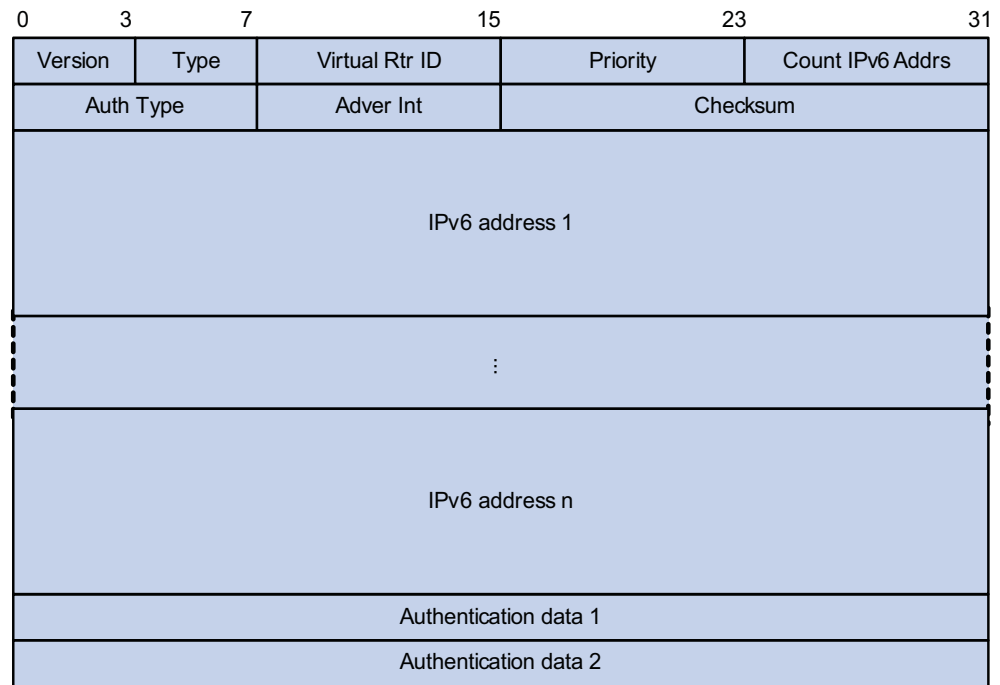


As shown in Figure 320, an IPv4-based VRRP packet consists of the following fields:

- Version: Version number of the protocol, 2 for VRRPv2.
- Type: Type of the VRRP packet. Only one VRRP packet type is present, that is, VRRP advertisement, which is represented by 1.
- Virtual Rtr ID (VRID): Number of the virtual router, that is, number of the standby group. It ranges from 1 to 255.
- Priority: Priority of the switch in the standby group, in the range 0 to 255. A greater value represents a higher priority.
- Count IP Adrs: Number of virtual IP addresses for the standby group. A standby group can have multiple virtual IP addresses.
- Auth Type: Authentication type. 0 means no authentication, 1 means simple authentication, and 2 means MD5 authentication.
- Adver Int: Interval for sending advertisement packets, in seconds. The default is 1.
- Checksum: 16-bit checksum for validating the data in VRRP packets.
- IP Address: Virtual IP address entry of the standby group. The allowed number is given by the Count IP Adrs field.
- Authentication Data: Authentication key. Currently, this field is used only for simple authentication and is 0 for any other authentication modes.

IPv6-based VRRP packet format

Figure 321 IPv6-based VRRP packet format



As shown in Figure 321, an IPv6-based VRRP packet consists of the following fields:

- Version: Version number of the protocol, 3 for VRRPv3.
- Type: Type of the VRRP packet. Only one VRRP packet type is present, that is, VRRP advertisement, which is represented by 1.
- Virtual Rtr ID (VRID): Number of the virtual router, that is, number of the standby group. It ranges from 1 to 255.
- Priority: Priority of the switch in the standby group, in the range 0 to 255. A greater value represents a higher priority.
- Count IPv6 Addr: Number of virtual IPv6 addresses for the standby group. A standby group can have multiple virtual IPv6 addresses.
- Auth Type: Authentication type. 0 means no authentication, 1 means simple authentication. VRRPv3 does not support MD5 authentication.
- Adver Int: Interval for sending advertisement packets, in centiseconds. The default is 100.
- Checksum: 16-bit checksum for validating the data in VRRPv3 packets.
- IPv6 Address: Virtual IPv6 address entry of the standby group. The allowed number is given by the Count IPv6 Addr field.
- Authentication Data: Authentication key. Currently, this field is used only for simple authentication and is 0 for any other authentication modes.

Principles of VRRP

- With VRRP enabled, the switches determine their respective roles in the standby group by priority. The switch with the highest priority becomes the

master, while the others are the backups. The master sends VRRP advertisement packets periodically to notify the backups that it is working properly, and each of the backups starts a timer to wait for advertisement packets from the master.

- In preemption mode, when a backup receives a VRRP advertisement packet, it compares the priority in the packet with that of its own. If its priority is higher, it becomes the master; otherwise, it remains a backup.
- In non-preemption mode, the switch in the standby group remains as a master or backup as long as the master does not fail. The backup will not become the master even if the former is configured with a higher priority.
- If the timer of a backup expires but the backup still does not receive any VRRP advertisement packet, it considers that the master fails. In this case, the backup switch considers itself as the master switch and sends VRRP advertisements to start the election process to elect a new master switch for forwarding packets.

VRRP Interface Tracking

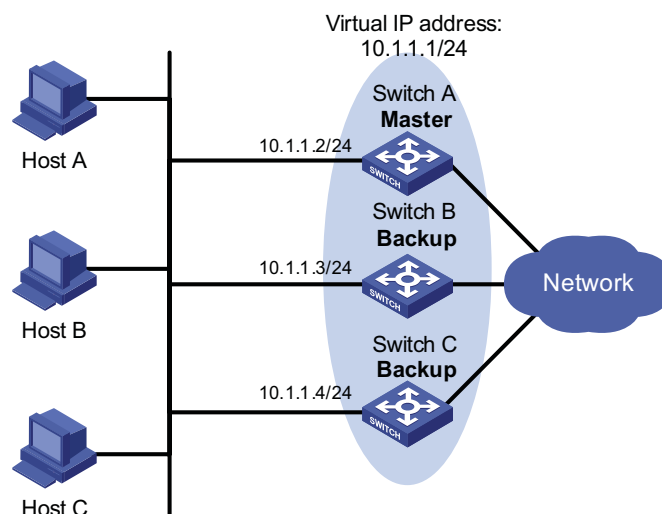
The interface tracking function expands the backup functionality of VRRP. It provides backup not only when the interface to which a standby group is assigned fails but also when other interfaces on the switch become unavailable. This is achieved by tracking interfaces. When a monitored interface goes down, the priority of the switch owning the interface is automatically decreased by a specified value, allowing a higher priority switch in the standby group to become the master.

VRRP Application (Taking IPv4-Based VRRP for Example)

Master/backup

In master/backup mode, only one switch, the master, provides services. When the master fails, a new master is elected from the original backups. This mode requires only one standby group, in which each switch holds different priorities and the one with the highest priority becomes the master, as shown in Figure 322.

Figure 322 VRRP in master/backup mode



At the beginning, Switch A is the master and therefore can forward packets to external networks, while Switch B and Switch C are backups and are thus in the state of listening. If Switch A fails, Switch B and Switch C will elect for the new

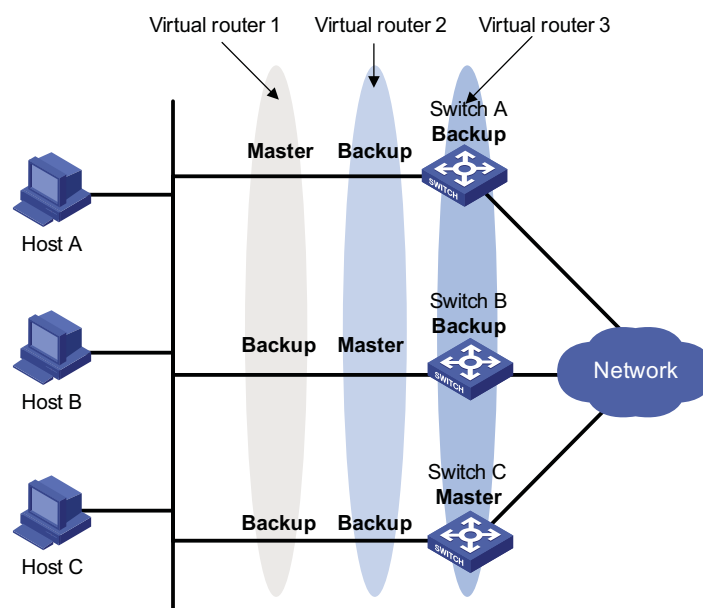
master. The new master takes over the forwarding task to provide services to hosts on the LAN.

Load balancing

You can create more than one standby group on an interface of a switch, allowing the switch to be the master of one standby group but a backup of another at the same time.

In load balancing mode, multiple switches provide services at the same time. This mode requires two or more standby groups, each of which includes a master and one or more backups. The masters of the standby groups can be assumed by different switches, as shown in Figure 323.

Figure 323 VRRP in load balancing mode



A switch can be in multiple standby groups and hold a different priority in different group.

In Figure 323, three standby groups are present:

- Standby group 1: Switch A is the master; Switch B and Switch C are the backups.
- Standby group 2: Switch B is the master; Switch A and Switch C are the backups.
- Standby group 3: Switch C is the master; Switch A and Switch B are the backups.

For load balancing among Switch A, Switch B, and Switch C, hosts on the LAN need to be configured to use standby group 1, 2, and 3 as the default gateways respectively. When configuring VRRP priorities, ensure that each switch holds such a priority in each standby group that it will take the expected role in the group.

Configuring VRRP for IPv4

VRRP for IPv4 Configuration Task List

Complete these tasks to configure VRRP for IPv4:

Task	Remarks
"Enabling Users to Ping Virtual IP Addresses" on page 1081	Optional
"Configuring the Association Between Virtual IP Address and MAC Address" on page 1081	Optional
"Creating Standby Group and Configuring Virtual IP Address" on page 1082	Required
"Configuring Standby Group Priority, Preemption Mode and Interface Tracking" on page 1083	Optional
"Configuring VRRP Packet Attributes" on page 1083	Optional

Enabling Users to Ping Virtual IP Addresses

You can configure whether the master switch responds to the received ICMP echo requests, that is, whether the virtual IP address of a standby group can be successfully pinged.

Follow these steps to enable a user to successfully ping the virtual IP addresses of standby groups:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable users to ping virtual IP address of the standby group	vrrp ping-enable	Optional Enabled by default.



CAUTION: Configure this function before creating a standby group. Otherwise, your configuration will fail.

Configuring the Association Between Virtual IP Address and MAC Address

After the virtual IP address of a standup group is associated with a MAC address, the master switch takes the configured MAC address as the source MAC address of the packets to be sent, so that the hosts in the internal network can learn the association between the IP address and the MAC address and thus forward the packets to be forwarded to the other network segments to the master switch properly.

There are two types of association between virtual IP address and MAC address:

- Virtual IP address is associated with virtual router MAC address

By default, a MAC address is created for a standby group after the standby group is created, and the virtual IP address is associated with the virtual MAC address. With such association adopted, the hosts in the internal network need not update the association between IP address and MAC address when the master switch changes.

- Virtual IP address is associated with real MAC address of the interface

When an IP address owner exists in a standby group, if you associate the virtual IP address with the virtual MAC address, two MAC addresses are associated with an

IP address. In this case, you can associate the virtual IP address of the standby group with the real MAC address, so that the packets from a host are forwarded to the IP address owner according to the real MAC address.

Follow these steps to configure the association between MAC address and virtual IP address:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the association between virtual IP address and MAC address	vrrp method { real-mac virtual-mac }	Optional The virtual MAC address is associated with the virtual IP address by default.



CAUTION: You should configure this function before creating a standby group. Otherwise, you cannot modify the mapping between the virtual IP address and the MAC address.

Creating Standby Group and Configuring Virtual IP Address

You need to configure a virtual IP address for a standby group when creating the standby group. A VRRP standby group is created automatically when you specify the first virtual IP address for the standby group. If you specify a virtual IP address for the standby group later, the virtual IP address is only added to the virtual IP address list of the VRRP standby group.

Configuration prerequisites

Before creating standby group and configuring virtual IP address, you should first configure the IP address of the interface and ensure that the virtual IP address to be configured is in the same network segment as the IP address of the interface.

Configuration procedure

Follow these steps to create standby group and configure virtual IP address:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter the specified interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Create standby group and configure virtual IP address of the standby group	vrrp vrid <i>virtual-router-id</i> virtual-ip <i>virtual-address</i>	Required Standup group is not created by default.



CAUTION:

- The maximum number of standby groups on an interface and the maximum number of virtual IP addresses in a standby group vary by device.
- A standby group is removed after you remove all the virtual IP addresses in it. In addition, configurations on that standby group no longer take effect.
- The virtual IP address of the virtual router can be either an unused IP address on the segment where the standby group resides or the IP address of an interface on a switch in the standby group. In the latter case, the switch is called the IP address owner.

- The virtual IP address of the standby group cannot be 0.0.0.0, 255.255.255.255, loopback address, non A/B/C address and other illegal IP addresses such as 0.0.0.1.
- Only when the configured virtual IP address and the interface IP address belong to the same segment and are legal host addresses can the standby group operate normally. If the configured virtual IP address and the interface IP address do not belong to the same network segment, or the configured IP address is the network address or network broadcast address of the network segment that the interface IP address belongs to, the state of the standby group is always **initialize** though you can perform the configuration successfully, that is, VRRP does not take effect in this case.

Configuring Standby Group Priority, Preemption Mode and Interface Tracking

Configuration prerequisites

Before you configure these features, you should first create a standby group on the interface and configure virtual IP address for it.

Configuration procedure

By configuring switch priority, preemption mode and interface tracking, you can decide which switch in the standby group serves as the Master.

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure switch priority in the standby group	vrrp vrid <i>virtual-router-id</i> priority <i>priority-value</i>	Optional 100 by default.
Configure the switch in the standby group to work in preemption mode and configure preemption delay	vrrp vrid <i>virtual-router-id</i> preempt-mode [timer delay <i>delay-value</i>]	Optional The switch in the standby group works in preemption mode and the preemption delay is 0 seconds by default.
Configure the interface to be tracked	vrrp vrid <i>virtual-router-id</i> track interface <i>interface-type</i> <i>interface-number</i> [reduced <i>priority-reduced</i>]	Optional No interface is being tracked by default.



CAUTION:

- The priority of an IP address owner is always 255 and not configurable.
- Interface tracking is not configurable to an IP address owner.
- The priority of a device is restored if the state of the interface under tracking changes from down to up.

Configuring VRRP Packet Attributes

Configuration prerequisites

Before configuring the relevant attributes of VRRP packets, you should first create the standby group and configure the virtual IP address.

Configuration procedure

Follow these steps to configure VRRP packet attributes:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter the specified interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure the authentication mode and authentication key when the standby groups send and receive VRRP packets	vrrp vrid <i>virtual-router-id</i> authentication-mode { md5 simple } <i>key</i>	Optional Authentication is not performed by default
Configure the time interval for the Master in the standby group to send VRRP advertisement	vrrp vrid <i>virtual-router-id</i> timer advertise <i>adver-interval</i>	Optional 1 second by default
Disable TTL check on VRRP packets	vrrp un-check ttl	Optional Enabled by default Do not create a standby group before executing this command.



- You may configure different authentication modes and authentication keys for the standby groups on an interface. However, the members of the same standby group must use the same authentication mode and authentication key.
- Factors like excessive traffic or different timer setting on switches can cause the Backup timer to time-out abnormally and trigger a change of the state. To solve this problem, you can prolong the time interval to send VRRP packets and configure a preemption delay.

Displaying and Maintaining VRRP for IPv4

To do...	Use the command...	Remarks
Display VRRP status	display vrrp [verbose] [interface <i>interface-type</i> <i>interface-number</i> [vrid <i>virtual-router-id</i>]]	Available in any view
Display VRRP statistics	display vrrp statistics [interface <i>interface-type</i> <i>interface-number</i> [vrid <i>virtual-router-id</i>]]	Available in any view
Remove VRRP statistics	reset vrrp statistics [interface <i>interface-type</i> <i>interface-number</i> [vrid <i>virtual-router-id</i>]]	Available in user view

Configuring VRRP for IPv6

VRRP for IPv6 Configuration Task List

Complete these tasks to configure VRRP for IPv6:

Task	Remarks
"Enabling Users to Ping Virtual IPv6 Addresses" on page 1085	Optional
"Configuring the Association Between Virtual IPv6 Address and MAC Address" on page 1085	Optional

Task	Remarks
"Creating Standby Group and Configuring Virtual IPv6 Address" on page 1086	Required
"Configuring Standby Group Priority, Preemption Mode and Interface Tracking" on page 1087	Optional
"Configuring VRRP Packet Attributes" on page 1087	Optional

Enabling Users to Ping Virtual IPv6 Addresses

You can configure whether the master switch responds to the received ICMPv6 echo requests, that is, whether the virtual IPv6 address of a standby group can be pinged through.

Follow these steps to enable a user to successfully ping the virtual IPv6 addresses of standby groups:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable a user to ping virtual IPv6 address of the standby group	vrrp ipv6 ping-enable	Optional Enabled by default



CAUTION: You should configure this function before creating a standby group. Otherwise, you cannot ping the virtual IPv6 addresses of standby groups.

Configuring the Association Between Virtual IPv6 Address and MAC Address

After the virtual IPv6 address of a standup group is associated with the MAC address, the master switch takes the configured MAC address as the source MAC address of the packets to be sent, so that the hosts in the internal network can learn the association between the IPv6 address and the MAC address and thus forward the packets to be forwarded to the other network segments to the master switch properly.

There are two types of association between virtual IPv6 address and MAC address:

- Virtual IPv6 address is associated with virtual router MAC address

By default, a MAC address is created for a standby group after the standby group is created, and the virtual IPv6 address is associated with the virtual MAC address. With such association adopted, the hosts in the internal network need not update the association between IPv6 address and MAC address when the master switch changes.

- Virtual IPv6 address is associated with real MAC address of the interface

When an IP address owner exists in a standby group, if you associate the virtual IPv6 address with the virtual MAC address, two MAC addresses are associated with an IPv6 address. In this case, you can associate the virtual IPv6 address of the standby group with the real MAC address, so that the packets from a host is forwarded to the IP address owner according the real MAC address.

Follow these steps to configure the association between MAC address and virtual IPv6 address:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the association between virtual IPv6 address and MAC address	vrrp ipv6 method { real-mac virtual-mac }	Optional The virtual MAC address of the standby group is associated with the virtual IPv6 address by default.



CAUTION: You should configure this function before creating a standby group. Otherwise, you cannot modify the mapping between the virtual IPv6 address and the MAC address.

Creating Standby Group and Configuring Virtual IPv6 Address

You need to configure a virtual IPv6 address for a standby group when creating the standby group. A VRRP standby group is created automatically when you specify the first virtual IPv6 address for the standby group. If you specify a virtual IPv6 address for the standby group later, the virtual IPv6 address is only added to the virtual IPv6 address list of the VRRP standby group.

Configuration prerequisites

Before creating standby group and configuring virtual IPv6 address, you should first configure the IPv6 address of the interface and ensure that the virtual IPv6 address to be configured is in the same network segment as the IPv6 address of the interface.

Configuration procedure

Follow these steps to create standby group and configure its virtual IPv6 address:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter the specified interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Create standby group and configure its virtual IPv6 address	vrrp ipv6 vrid <i>virtual-router-id</i> virtual-ip <i>virtual-address</i> [link-local]	Required No standby group is created by default. The first virtual IPv6 address of the standby group must be a link local address. Only one link local address is allowed in a standby group, and must be removed the last.



CAUTION:

- The maximum number of standby groups on an interface and the maximum number of virtual IPv6 addresses in a standby group vary by device.
- A standby group is removed after you remove all the virtual IPv6 addresses in it. In addition, configurations on that standby group no longer take effect.

Configuring Standby Group Priority, Preemption Mode and Interface Tracking

Configuration prerequisites

Before configuring these features, you should first create the standby group and configure the virtual IPv6 address.

Configuration procedure

By configuring standby group priority, preemption mode and interface tracking, you can decide which switch in the standby group serves as the Master.

Follow these steps to configure standby group priority, preemption mode and interface tracking:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter the specified interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure the priority of the switch in the standby group	vrrp ipv6 vrid <i>virtual-router-id</i> priority <i>priority-value</i>	Optional 100 by default
Configure the switch in the standby to work in preemption mode and configure preemption delay of the standby group	vrrp ipv6 vrid <i>virtual-router-id</i> preempt-mode [timer delay <i>delay-value</i>]	Optional The switch in the standby group works in preemption mode and the preemption delay is zero seconds by default.
Configure the interface to be tracked	vrrp ipv6 vrid <i>virtual-router-id</i> track interface <i>interface-type</i> <i>interface-number</i> [reduced <i>priority-reduced</i>]	Optional No interface is being tracked by default.



CAUTION:

- *The priority of an IP address owner is always 255 and not configurable.*
- *Interface tracking is not configurable on an IP address owner.*
- *The priority of a device is reset if the state of the interface under tracking changes from down to up.*

Configuring VRRP Packet Attributes

Configuration prerequisites

Before configuring the relevant attributes of VRRP packets, you should first create the standby group and configure the virtual IPv6 address.

Configuration procedure

Follow these steps to configure VRRP packet attributes:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter the specified interface view	interface <i>interface-type</i> <i>interface-number</i>	-

To do...	Use the command...	Remarks
Configure the authentication mode and authentication key when the standby groups send and transmit VRRP packets	vrrp ipv6 vrid <i>virtual-router-id</i> authentication-mode simple <i>key</i>	Optional Authentication is not performed by default
Configure the time interval for the Master in the standby group to send VRRP advertisement	vrrp ipv6 vrid <i>virtual-router-id</i> timer advertise <i>adver-interval</i>	Optional 100 centiseconds by default

You may configure different authentication modes and authentication keys for the standby groups on an interface. However, the members of the same standby group must use the same authentication mode and authentication key.

Factors like excessive traffic or different timer setting on switches can cause the Backup timer to time-out abnormally and change the state. To solve this problem, you can prolong the time interval to send VRRP packets and configure a delay for preemption.

Displaying and Maintaining VRRP for IPv6

To do...	Use the command...	Remarks
Display VRRP status	display vrrp ipv6 [verbose] [interface <i>interface-type interface-number [vrid</i> <i>virtual-router-id]]</i>	Available in any view
Display VRRP statistics	display vrrp ipv6 statistics [interface <i>interface-type interface-number [vrid</i> <i>virtual-router-id]]</i>	Available in any view
Remove VRRP statistics	reset vrrp ipv6 statistics [interface <i>interface-type interface-number [vrid</i> <i>virtual-router-id]]</i>	Available in user view

IPv4-Based VRRP Configuration Examples

This section provides these configuration examples:

- “Single VRRP Standby Group Configuration Example” on page 1088
- “VRRP Interface Tracking Configuration Example” on page 1091
- “Multiple VRRP Standby Group Configuration Example” on page 1094

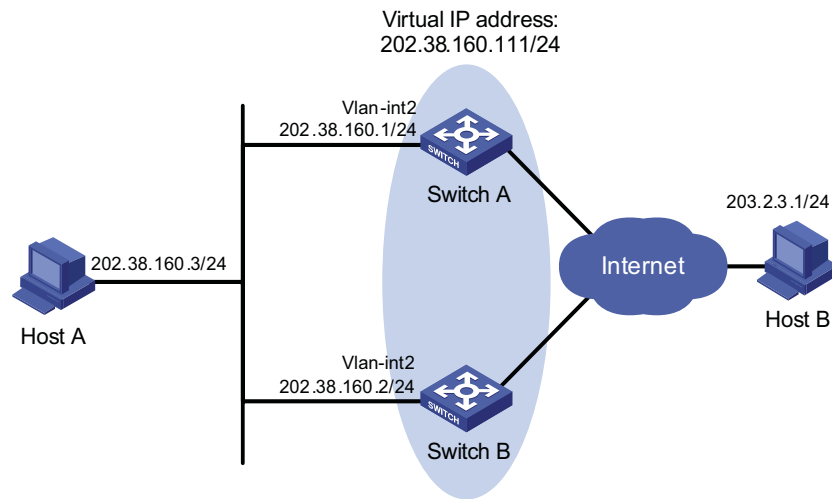
Single VRRP Standby Group Configuration Example

Network requirements

- Host A needs to access Host B on the Internet, using 202.38.160.111/24 as its default gateway.
- Switch A and Switch B belong to standby group 1 with the virtual IP address of 202.38.160.111/24.
- If Switch A operates normally, packets sent from Host A to Host B are forwarded by Switch A; if Switch A fails, packets sent from Host A to Host B are forwarded by Switch B.

Network diagram

Figure 324 Network diagram for single VRRP standby group configuration



Configuration procedure

1 Configure Switch A

Configure VLAN 2.

```
<SwitchA> system-view
[SwitchA] vlan 2
[SwitchA-vlan2] port GigabitEthernet 1/0/5
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 202.38.160.1 255.255.255.0
```

Create standby group 1 and set its virtual IP address to be 202.38.160.111.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.111
```

Set the priority of Switch A in standby group 1 to 110.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
```

Set Switch A to work in preemption mode. The preemption delay is five seconds.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 preempt-mode timer delay 5
```

2 Configure Switch B

Configure VLAN 2.

```
<SwitchB> system-view
[SwitchB] vlan 2
[SwitchB-Vlan2] port GigabitEthernet 1/0/5
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 202.38.160.2 255.255.255.0
```

Create standby group 1 and set its virtual IP address to be 202.38.160.111.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.111
```

Set Switch B to work in preemption mode. The preemption delay is five seconds.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 preempt-mode timer delay 5
```

3 Verify the configuration

After the configuration, Host B can be pinged through on Host A. You can use the **display vrrp** command to verify the configuration.

Display detailed information of standby group 1 on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp verbose
IPv4 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri      : 110
Preempt Mode    : YES
Auth Type       : NONE
Virtual IP      : 202.38.160.111
Virtual MAC     : 0000-5e00-0101
Master IP       : 202.38.160.1
Adver. Timer    : 1
State           : Master
Run Pri         : 110
Delay Time      : 5
```

Display detailed information of standby group 1 on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp verbose
IPv4 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri      : 100
Preempt Mode    : YES
Auth Type       : NONE
Virtual IP      : 202.38.160.111
Master IP       : 202.38.160.1
Adver. Timer    : 1
State           : Backup
Run Pri         : 100
Delay Time      : 5
```

The above information indicates that in standby group 1 Switch A is the master, Switch B is the backup and packets sent from Host A to Host B are forwarded by Switch A.

If Switch A fails, you can still ping through Host B on Host A. Use the **display vrrp** command to view the detailed information of the standby group on Switch B.

If Switch A fails, the detailed information of standby group 1 on Switch B is displayed.

```
[SwitchB-Vlan-interface2] display vrrp verbose
IPv4 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
```

```

VRID                : 1                      Adver. Timer       : 1
Admin Status       : UP                      State              : Master
Config Pri         : 100                     Run Pri            : 100
Preempt Mode       : YES                     Delay Time         : 5
Auth Type          : NONE
Virtual IP         : 202.38.160.111
Virtual MAC        : 0000-5e00-0101
Master IP          : 202.38.160.2

```

The above information indicates that if Switch A fails, Switch B becomes the master, and packets sent from Host A to Host B are forwarded by Switch B.

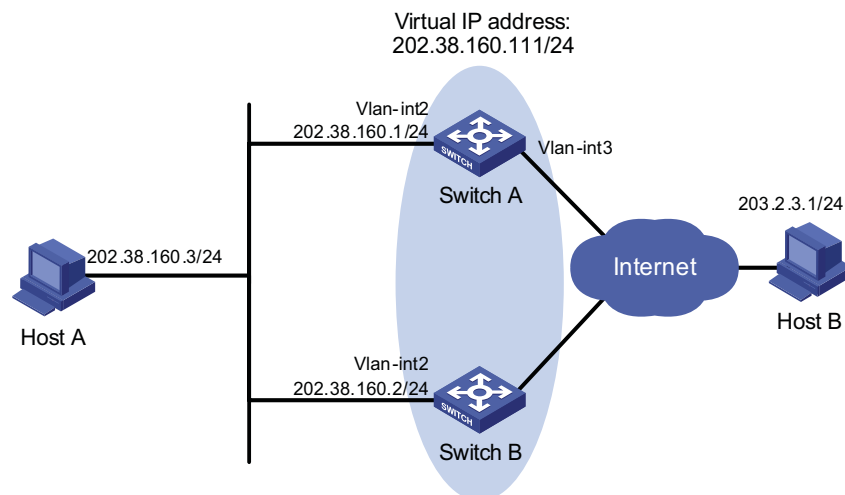
VRRP Interface Tracking Configuration Example

Network requirements

- Host A needs to access Host B on the Internet, using 202.38.160.111/24 as its default gateway.
- Switch A and Switch B belong to standby group 1 with the virtual IP address of 202.38.160.111.
- If Switch A operates normally, packets sent from Host A to Host B are forwarded by Switch A; if Switch A is in work, but its VLAN-interface 3 which connects to the Internet is not available, packets sent from Host A to Host B are forwarded by Switch B.

Network diagram

Figure 325 Network diagram for VRRP interface tracking



Configuration procedure

1 Configure Switch A

Configure VLAN 2.

```

<SwitchA> system-view
[SwitchA] vlan 2
[SwitchA-vlan2] port GigabitEthernet 1/0/5
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 202.38.160.1 255.255.255.0

```

Create a standby group 1 and set its virtual IP address to 202.38.160.111.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.111
```

Configure the priority of Switch A in the standby group to 110.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
```

Configure the authentication mode of the standby group to **simple** and authentication key to **hello**.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 authentication-mode simple hello
```

Set the interval for Master to send VRRP advertisement to five seconds.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 timer advertise 5
```

Set the interface to be tracked.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 track interface vlan-interface 3 reduced 30
```

2 Configure Switch B

Configure VLAN 2.

```
<SwitchB> system-view
[SwitchB] vlan 2
[SwitchB-vlan2] port GigabitEthernet 1/0/5
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 202.38.160.2 255.255.255.0
```

Create a standby group 1 and set its virtual IP address to 202.38.160.111.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.111
```

Configure the authentication mode of the standby group to **simple** and authentication key to **hello**.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 authentication-mode simple hello
```

Set the interval for Master to send VRRP advertisement to five seconds.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 timer advertise 5
```

3 Verify the configuration

After the configuration, Host B can be pinged through on Host A. You can use the **display vrrp** command to verify the configuration.

Display detailed information of standby group 1 on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp verbose
IPv4 Standby Information:
Run Method          : VIRTUAL-MAC
```



```

Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID           : 1                               Adver. Timer   : 5
Admin Status   : UP                             State          : Master
Config Pri     : 110                            Run Pri        : 110
Preempt Mode   : YES                            Delay Time     : 0
Auth Type      : SIMPLE TEXT                     Key            : hello
Track IF       : Vlan-interface3                 Pri Reduced    : 30
Virtual IP     : 202.38.160.111
Virtual MAC    : 0000-5e00-0101
Master IP      : 202.38.160.1

```

Display detailed information of standby group 1 on Switch B.

```

[SwitchB-Vlan-interface2] display vrrp verbose
IPv4 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID           : 1                               Adver. Timer   : 5
Admin Status   : UP                             State          : Backup
Config Pri     : 100                            Run Pri        : 100
Preempt Mode   : YES                            Delay Time     : 0
Auth Type      : SIMPLE TEXT                     Key            : hello
Virtual IP     : 202.38.160.111
Master IP      : 202.38.160.1

```

The above information indicates that in standby group 1 Switch A is the master, Switch B is the backup and packets sent from Host A to Host B are forwarded by Switch A.

If Switch A is in work, but when its interface VLAN-interface 3 that connects to the Internet is not available, you can still ping through Host B on Host A. Use the **display vrrp** command to view the detailed information of the standby group.

If VLAN-interface 3 on Switch A is not available, the detailed information of standby group 1 on Switch A is displayed.

```

[SwitchA-Vlan-interface2] display vrrp verbose
IPv4 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID           : 1                               Adver. Timer   : 5
Admin Status   : UP                             State          : Backup
Config Pri     : 110                            Run Pri        : 80
Preempt Mode   : YES                            Delay Time     : 0
Auth Type      : SIMPLE TEXT                     Key            : hello
Track IF       : Vlan-interface3                 Pri Reduced    : 30
Virtual IP     : 202.38.160.111
Master IP      : 202.38.160.2

```

If VLAN-interface 3 on Switch A is not available, the detailed information of standby group 1 on Switch B is displayed.

```

[SwitchB-Vlan-interface2] display vrrp verbose
IPv4 Standby Information:

```

```

Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri      : 100
Preempt Mode    : YES
Auth Type       : SIMPLE TEXT
Virtual IP      : 202.38.160.111
Virtual MAC     : 0000-5e00-0101
Master IP       : 202.38.160.2
Adver. Timer    : 5
State           : Master
Run Pri         : 100
Delay Time      : 0
Key             : hello
    
```

The above information indicates that if VLAN-interface 3 on Switch A is not available, the priority of Switch A is reduced to 80 and it becomes the backup. Switch B becomes the master and packets sent from Host A to Host B are forwarded by Switch B.

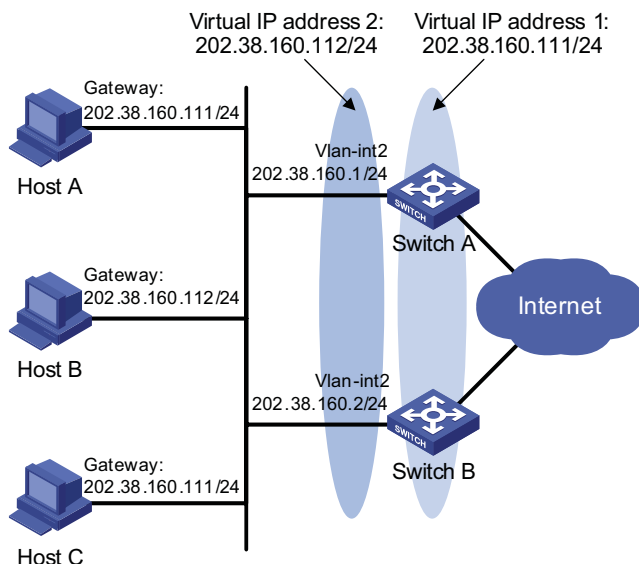
Multiple VRRP Standby Group Configuration Example

Network requirements

- In the segment 202.38.160.0/24, some hosts use 202.38.160.111/24 as their default gateway and some hosts use 202.38.160.112/24 as their default gateway.
- Load sharing and mutual backup between default gateways can be implemented by using VRRP standby groups.

Network diagram

Figure 326 Network diagram for multiple VRRP standby group configuration



Configuration procedure

- 1 Configure Switch A

Configure VLAN 2.

```

<SwitchA> system-view
[SwitchA] vlan 2
[SwitchA-vlan2] port GigabitEthernet 1/0/5
    
```

```
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ip address 202.38.160.1 255.255.255.0
```

Create a standby group 1 and set its virtual IP address to 202.38.160.111.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.111
```

Configure the priority of Switch A in standby group 1 to 110.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
```

Create a standby group 2 and set its virtual IP address to 202.38.160.112.

```
[SwitchA-Vlan-interface2] vrrp vrid 2 virtual-ip 202.38.160.112
```

2 Configure Switch B

Configure VLAN 2.

```
<SwitchB> system-view
[SwitchB] vlan 2
[SwitchB-vlan2] port GigabitEthernet 1/0/5
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ip address 202.38.160.2 255.255.255.0
```

Create a standby group 1 and set its virtual IP address to 202.38.160.111.

```
[SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 202.38.160.111
```

Create a standby group 2 and set its virtual IP address to 202.38.160.112.

```
[SwitchB-Vlan-interface2] vrrp vrid 2 virtual-ip 202.38.160.112
```

Configure the priority of Switch B in standby group 2 to 110.

```
[SwitchB-Vlan-interface2] vrrp vrid 2 priority 110
```

3 Verify the configuration

You can use the **display vrrp** command to verify the configuration.

Display detailed information of the standby group on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp verbose
IPv4 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri     : 110
Preempt Mode   : YES
Auth Type      : NONE
Virtual IP     : 202.38.160.111
Virtual MAC    : 0000-5e00-0101
Master IP     : 202.38.160.1
Adver. Timer   : 1
State          : Master
Run Pri        : 110
Delay Time     : 0
```

```

Interface      : Vlan-interface2
VRID           : 2
Admin Status   : UP
Config Pri     : 100
Preempt Mode   : YES
Auth Type      : NONE
Virtual IP     : 202.38.160.112
Master IP      : 202.38.160.2
Adver. Timer   : 1
State          : Backup
Run Pri        : 100
Delay Time     : 0

```

Display detailed information of the standby group on Switch B.

```

[SwitchB-Vlan-interface2] display vrrp verbose
IPv4 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri      : 100
Preempt Mode    : YES
Auth Type       : NONE
Virtual IP      : 202.38.160.111
Master IP       : 202.38.160.1
Adver. Timer    : 1
State           : Backup
Run Pri         : 100
Delay Time      : 0

Interface       : Vlan-interface2
VRID            : 2
Admin Status    : UP
Config Pri      : 110
Preempt Mode    : YES
Auth Type       : NONE
Virtual IP      : 202.38.160.112
Virtual MAC     : 0000-5e00-0102
Master IP       : 202.38.160.2
Adver. Timer    : 1
State           : Master
Run Pri         : 110
Delay Time      : 0

```

The above information indicates that in standby group 1 Switch A is the master, Switch B is the backup and the host with the default gateway of 202.38.160.111/24 accesses the Internet through Switch A; in standby group 2 Switch A is the backup, Switch B is the master and the host with the default gateway of 202.38.160.112/24 accesses the Internet through Switch B.

IPv6-Based VRRP Configuration Examples

This section provides these configuration examples:

- "Single VRRP Standby Group Configuration Example" on page 1096
- "VRRP Interface Tracking Configuration Example" on page 1099
- "Multiple VRRP Standby Group Configuration Example" on page 1102

Single VRRP Standby Group Configuration Example

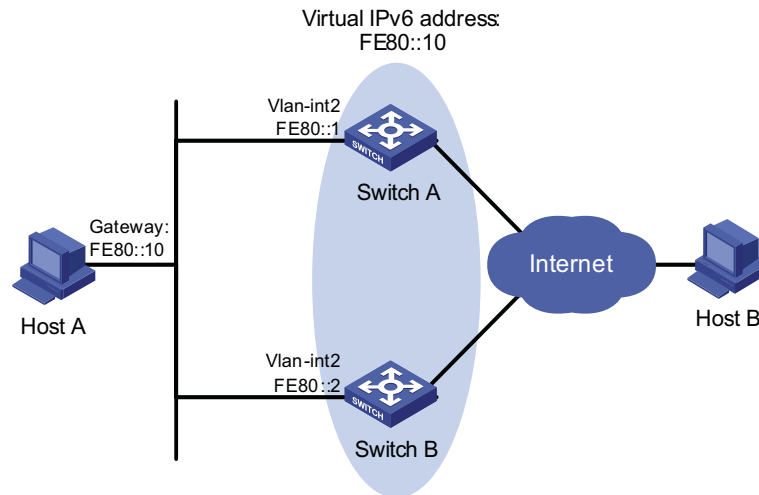
Network requirements

- Host A needs to access Host B on the Internet, using FE80::10 as its default gateway.
- Switch A and Switch B belong to standby group 1 with the virtual IP address of FE80::10.

- If Switch A operates normally, packets sent from Host A to Host B are forwarded by Switch A; if Switch A fails, packets sent from Host A to Host B are forwarded by Switch B.

Network diagram

Figure 327 Network diagram for single VRRP standby group configuration



Configuration procedure

1 Configure Switch A

Configure VLAN 2.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] vlan 2
[SwitchA-vlan2] port GigabitEthernet 1/0/5
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ipv6 address fe80::1 link-local
[SwitchA-Vlan-interface2] ipv6 address 1::1 64
```

Create a standby group 1 and set its virtual IP address to FE80::10.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
```

Set the priority of Switch A in standby group 1 to 110.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 priority 110
```

Set Switch A to work in preemption mode.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 preempt-mode
```

Enable Switch A to send RA messages.

```
[SwitchA-Vlan-interface2] undo ipv6 nd ra halt
```

2 Configure Switch B

Configure VLAN 2.

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] vlan 2
[SwitchB-vlan2] port GigabitEthernet 1/0/5
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ipv6 address fe80::2 link-local
```

Create a standby group 1 and set its virtual IP address to FE80::10.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
```

Enable Switch B to send RA messages.

```
[SwitchB-Vlan-interface2] undo ipv6 nd ra halt
```

3 Verify the configuration

After the configuration, Host B can be pinged through on Host A. You can use the **display vrrp ipv6** command to verify the configuration.

Display detailed information of standby group 1 on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp ipv6 verbose
IPv6 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri     : 110
Preempt Mode    : YES
Auth Type       : NONE
Virtual IP      : FE80::10
Virtual MAC     : 0000-5e00-0201
Master IP       : FE80::1
Adver. Timer    : 100
State           : Master
Run Pri         : 110
Delay Time      : 0
```

Display detailed information of standby group 1 on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp ipv6 verbose
IPv6 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri     : 100
Preempt Mode    : YES
Auth Type       : NONE
Virtual IP      : FE80::10
Master IP       : FE80::1
Adver. Timer    : 100
State           : Backup
Run Pri         : 100
Delay Time      : 0
```

The above information indicates that in standby group 1 Switch A is the master, Switch B is the backup and packets sent from Host A to Host B are forwarded by Switch A.

If Switch A fails, you can still ping through Host B on Host A. You can use the **display vrrp ipv6** command to view the detailed information of the standby group on Switch B.

If Switch A fails, the detailed information of standby group 1 on Switch B is displayed.

```
[SwitchB-Vlan-interface2] display vrrp ipv6 verbose
IPv6 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri      : 100
Preempt Mode    : YES
Auth Type       : NONE
Virtual IP      : FE80::10
Virtual MAC     : 0000-5e00-0201
Master IP       : FE80::2
Adver. Timer    : 100
State           : Master
Run Pri         : 100
Delay Time      : 0
```

The above information indicates that if Switch A fails, Switch B becomes the master, and packets sent from Host A to Host B are forwarded by Switch B.

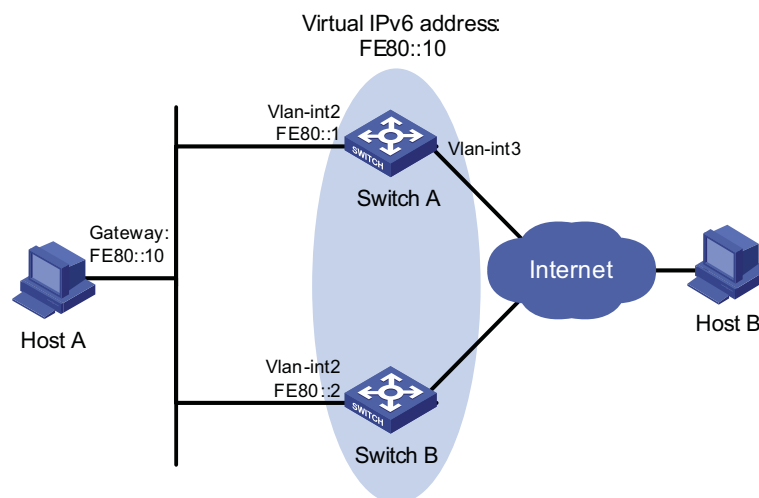
VRRP Interface Tracking Configuration Example

Network requirements

- Host A needs to access Host B on the Internet, using FE80::10 as its default gateway.
- Switch A and Switch B belong to standby group 1 with the virtual IP address of FE80::10.
- If Switch A operates normally, packets sent from Host A to Host B are forwarded by Switch A; if Switch A is in work, but its VLAN-interface 3 which connects to the Internet is not available, packets sent from Host A to Host B are forwarded by Switch B.

Network diagram

Figure 328 Network diagram for VRRP interface tracking



Configuration procedure

1 Configure Switch A

Configure VLAN 2.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] vlan 2
[SwitchA-vlan2] port GigabitEthernet 1/0/5
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ipv6 address fe80::1 link-local
[SwitchA-Vlan-interface2] ipv6 address 1::1 64
```

Create a standby group 1 and set its virtual IP address to FE80::10.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
```

Set the priority of Switch A in standby group 1 to 110.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 priority 110
```

Set the authentication mode for standby group 1 to **simple** and authentication key to **hello**.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 authentication-mode simple hello
```

Set the VRRP advertisement interval to 500 centiseconds.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 timer advertise 500
```

Set Switch A work in preemption mode. The preemption delay is five seconds.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 preempt-mode timer delay 5
```

Set the interface to be tracked.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 track interface vlan-interface 3 reduced 30
```

2 Configure Switch B

Configure VLAN 2.

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB] vlan 2
[SwitchB-vlan2] port GigabitEthernet 1/0/5
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ipv6 address fe80::2 link-local
[SwitchB-Vlan-interface2] ipv6 address 1::2 64
```

Create a standby group 1 and set its virtual IP address to FE80::10.


```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-
local
```

Set the authentication mode for standby group 1 to **simple** and authentication key to **hello**.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 authentication-mode simpl
e hello
```

Set the VRRP advertisement interval to 500 centiseconds.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 timer advertise 500
```

Set Switch B to work in preemption mode. The preemption delay is five seconds.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 preempt-mode timer delay
5
```

3 Verify the configuration

After the configuration, Host B can be pinged through on Host A. You can use the **display vrrp ipv6** command to verify the configuration.

Display detailed information of standby group 1 on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp ipv6 verbose
IPv6 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri      : 110
Preempt Mode    : YES
Auth Type       : SIMPLE TEXT
Track IF        : Vlan-interface3
Virtual IP      : FE80::10
Virtual MAC     : 0000-5e00-0201
Master IP       : FE80::1
Adver. Timer    : 500
State           : Master
Run Pri         : 110
Delay Time      : 5
Key             : hello
Pri Reduced     : 30
```

Display detailed information of standby group 1 on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp ipv6 verbose
IPv6 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri      : 100
Preempt Mode    : YES
Auth Type       : SIMPLE TEXT
Track IF        : Vlan-interface3
Virtual IP      : FE80::10
Virtual MAC     : 0000-5e00-0201
Master IP       : FE80::1
Adver. Timer    : 500
State           : Backup
Run Pri         : 100
Delay Time      : 5
Key             : hello
```

The above information indicates that in standby group 1 Switch A is the master, Switch B is the backup and packets sent from Host A to Host B are forwarded by Switch A.

If Switch A is in work, but its interface VLAN-interface 3 is not available, you can still ping through Host B on Host A. You can use the **display vrrp ipv6** command to view the detailed information of the standby group.

If Switch A is in work, but its interface VLAN-interface 3 is not available, the detailed information of standby group 1 on Switch A is displayed.

```
[SwitchA-Vlan-interface2] display vrrp ipv6 verbose
IPv6 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri      : 110
Preempt Mode    : YES
Auth Type       : SIMPLE TEXT
Track IF        : Vlan-interface3
Virtual IP      : FE80::10
Master IP       : FE80::2
Adver. Timer    : 500
State           : Backup
Run Pri         : 80
Delay Time      : 5
Key             : hello
Pri Reduced     : 30
```

If Switch A is in work, but its interface VLAN-interface 3 is not available, the detailed information of standby group 1 on Switch B is displayed.

```
[SwitchB-Vlan-interface2] display vrrp ipv6 verbose
IPv6 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri      : 100
Preempt Mode    : YES
Auth Type       : SIMPLE TEXT
Virtual IP      : FE80::10
Virtual MAC     : 0000-5e00-0201
Master IP       : FE80::2
Adver. Timer    : 500
State           : Master
Run Pri         : 100
Delay Time      : 5
Key             : hello
```

The above information indicates that if VLAN-interface 3 on Switch A is not available, the priority of Switch A reduces to 80 and it becomes the backup. Switch B becomes the master and packets sent from Host A to Host B are forwarded by Switch B.

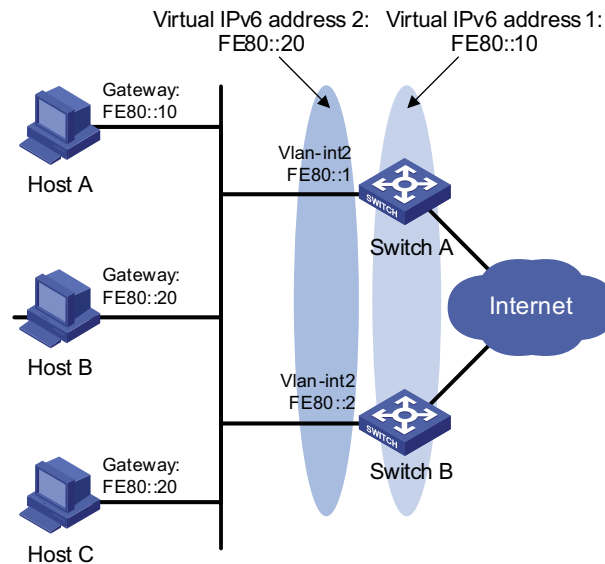
Multiple VRRP Standby Group Configuration Example

Network requirements

- In the network, some hosts use FE80::10 as their default gateway and some hosts use FE80::20 as their default gateway.
- Load sharing and mutual backup between default gateways can be implemented by using VRRP standby groups.

Network diagram

Figure 329 Network diagram for multiple VRRP standby group configuration



Configuration procedure

1 Configure Switch A

Configure VLAN 2.

```
<SwitchA> system-view
[SwitchA] ipv6
[SwitchA] vlan 2
[SwitchA-vlan2] port GigabitEthernet 1/0/5
[SwitchA-vlan2] quit
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] ipv6 address fe80::1 link-local
[SwitchA-Vlan-interface2] ipv6 address 1::1 64
```

Create standby group 1 and set its virtual IP address to FE80::10.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-local
```

Set the priority of Switch A in standby group 1 to 110.

```
[Switch-Vlan-interface2] vrrp ipv6 vrid 1 priority 110
```

Create standby group 2 and set its virtual IP address to FE80::20.

```
[SwitchA-Vlan-interface2] vrrp ipv6 vrid 2 virtual-ip fe80::20 link-local
```

2 Configure Switch B

Configure VLAN 2.

```
<SwitchB> system-view
[SwitchB] ipv6
[SwitchB-vlan2] port GigabitEthernet 1/0/5
```

```
[SwitchB-vlan2] quit
[SwitchB] interface vlan-interface 2
[SwitchB-Vlan-interface2] ipv6 address fe80::2 link-local
[SwitchB-Vlan-interface2] ipv6 address 1::2 64
```

Create standby group 1 and set its virtual IP address to FE80::10.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 1 virtual-ip fe80::10 link-
local
```

Create standby group 2 and set its virtual IP address to FE80::20.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 2 virtual-ip fe80::20 link-
local
```

Set the priority of Switch B in standby group 2 to 110.

```
[SwitchB-Vlan-interface2] vrrp ipv6 vrid 2 priority 110
```

3 Verify the configuration

You can use the **display vrrp ipv6** command to verify the configuration.

Display detailed information of the standby group on Switch A.

```
[SwitchA-Vlan-interface2] display vrrp ipv6 verbose
IPv6 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri      : 110
Preempt Mode    : YES
Auth Type       : NONE
Virtual IP      : FE80::10
Virtual MAC     : 0000-5e00-0201
Master IP       : FE80::1

Adver. Timer    : 100
State           : Master
Run Pri         : 110
Delay Time      : 0

Interface       : Vlan-interface2
VRID            : 2
Admin Status    : UP
Config Pri      : 100
Preempt Mode    : YES
Auth Type       : NONE
Virtual IP      : FE80::20
Master IP       : FE80::2

Adver. Timer    : 100
State           : Backup
Run Pri         : 100
Delay Time      : 0
```

Display detailed information of the standby group on Switch B.

```
[SwitchB-Vlan-interface2] display vrrp ipv6 verbose
IPv6 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri      : 100
Preempt Mode    : YES
Auth Type       : NONE
Virtual IP      : FE80::20
Master IP       : FE80::2

Adver. Timer    : 100
State           : Backup
Run Pri         : 100
Delay Time      : 0
```

```

Preempt Mode      : YES                      Delay Time       : 0
Auth Type         : NONE
Virtual IP        : FE80::10
Master IP         : FE80::1

Interface         : Vlan-interface2
VRID              : 2                        Adver. Timer     : 100
Admin Status     : UP                       State            : Master
Config Pri       : 110                      Run Pri         : 110
Preempt Mode     : YES                      Delay Time       : 0
Auth Type        : NONE
Virtual IP        : FE80::20
Virtual MAC       : 0000-5e00-0202
Master IP        : FE80::2

```

The above information indicates that in standby group 1 Switch A is the master, Switch B is the backup and the host with the default gateway of FE80::10 accesses the Internet through Switch A; in standby group 2 Switch A is the backup, Switch B is the master and the host with the default gateway of FE80::20 accesses the Internet through Switch B.



Multiple standby groups are commonly used in actual networking. In IPv6 network, you need to manually configure the default gateway for VRRP standby group to share load.

Troubleshooting VRRP Symptom 1:

The console screen displays error prompts frequently.

Analysis:

This error is probably due to the inconsistent configuration of the other switch in the standby group, or that a device is attempting to send illegitimate VRRP packets.

Solution:

- In the first case, modify the configuration.
- In the latter case, you have to resort to non-technical measures.

Symptom 2:

Multiple masters are present in the same standby group.

Analysis:

- If presence of multiple masters only lasts a short period, this is normal and requires no manual intervention.
- If it lasts long, you must ensure that these masters can receive VRRP packets and the packets received are legitimate.

Solution:

Ping between these masters, and do the following:

- If the ping fails, check network connectivity.
- If the ping succeeds, check that their configurations are consistent in terms of number of virtual IP addresses, virtual IP addresses, advertisement interval, and authentication.

Symptom 3:

Frequent VRRP state transition.

Analysis:

The VRRP advertisement interval is set too short.

Solution:

Increase the interval to sent VRRP advertisement or introduce a preemption delay

When configuring SSH, go to these sections for information you are interested in:

- “SSH2.0 Overview” on page 1107
- “Configuring the Device as an SSH Server” on page 1110
- “Configuring the Device as an SSH Client” on page 1115
- “Displaying and Maintaining SSH” on page 1118
- “SSH Server Configuration Examples” on page 1119
- “SSH Client Configuration Examples” on page 1125

SSH2.0 Overview

Secure Shell (SSH) offers an approach to securely logging into a remote device. By encryption and strong authentication, it protects devices against attacks such as IP spoofing and plain text password interception.

The device can not only work as an SSH server to support connections with SSH clients, but also work as an SSH client to allow users to establish SSH connections with a remote device acting as the SSH server.

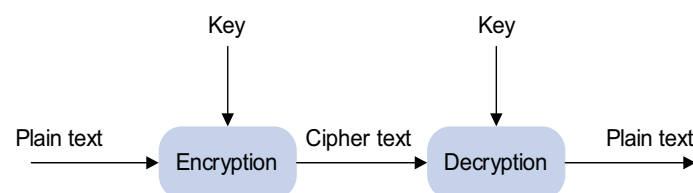


Currently, when acting as an SSH server, the device supports two SSH versions: SSH2 and SSH1. When acting as an SSH client, the device supports SSH2 only.

Algorithm and Key

Algorithm is a set of transformation rules for encryption and decryption. Information without being encrypted is known as plain text, while information that is encrypted is known as cipher text. Encryption and decryption are performed using a string of characters called a key, which controls the transformation between plain text and cipher text, for example, changing the plain text into cipher text or cipher text into plain text.

Figure 330 Encryption and decryption



Key-based algorithm is usually classified into symmetric key algorithm and asymmetric key algorithm.

Asymmetric Key Algorithm

Asymmetric key algorithm means that a key pair exists at both ends. The key pair consists of a private key and a public key. The public key is effective for both ends, while the private key is effective only for the local end.

Asymmetric key algorithm encrypts data using the public key and decrypts the data using the private key, thus ensuring data security.

You can also use the asymmetric key algorithm for digital signature. For example, user 1 adds his signature to the data using the private key, and then sends the data to user 2. User 2 verifies the signature using the public key of user 1. If the signature is correct, this means that the data originates from user 1.

Revest-Shamir-Adleman Algorithm (RSA) and Digital Signature Algorithm (DSA) are both asymmetric key algorithms. RSA can be used for data encryption and signature, whereas DSA is used for signatures only.



Currently, SSH2 supports both RSA and DSA.

SSH Operating Process

The session establishment between an SSH client and the SSH server involves the following five stages:

Table 88 Stages in establishing a session between the SSH client and the server

Stages	Description
"Version negotiation" on page 1108	SSH1 and SSH2 are supported. The two parties negotiate a version to use.
"Key and algorithm negotiation" on page 1109	SSH supports multiple algorithms. The two parties negotiate an algorithm for communication.
"Authentication" on page 1109	The SSH server authenticates the client in response to the client's authentication request.
"Session request" on page 1110	This client sends a session request to the server.
"Interactive session" on page 1110	The client and the server start to communicate with each other.

Version negotiation

- The server opens port 22 to listen to connection requests from clients.
- The client sends a TCP connection request to the server. After the TCP connection is established, the server sends the first packet to the client, which includes a version identification string in the format of "SSH-<primary protocol version number>.<secondary protocol version number>-<software version number>". The primary and secondary protocol version numbers constitute the protocol version number, while the software version number is used for debugging.
- The client receives and resolves the packet. If the protocol version of the server is lower but supportable, the client uses the protocol version of the server; otherwise, the client uses its own protocol version.
- The client sends to the server a packet that contains the number of the protocol version it decides to use. The server compares the version carried in the packet with that of its own to determine whether it can cooperate with the client.
- If the negotiation is successful, the server and the client proceed with key and algorithm negotiation; otherwise, the server breaks the TCP connection.



All the packets involved in the above steps are transferred in plain text.

Key and algorithm negotiation

- The server and the client send key algorithm negotiation packets to each other, which include the supported public key algorithm list, encryption algorithm list, MAC algorithm list, and compression algorithm list.
- Based on the received algorithm negotiation packets, the server and the client figure out the algorithms to be used.
- The server and the client use the DH key exchange algorithm and parameters such as the host key pair to generate the session key and session ID.

Through the above steps, the server and the client get the same session key, which is to be used to encrypt and decrypt data exchanged between the server and the client later. The server and the client use session ID in the authentication stage.



CAUTION: Before the negotiation, the server must have already generated the RSA and DSA key pairs, which are mainly used for generating the session key.

Authentication

- The client sends to the server an authentication request, which includes the username, authentication method and information related to the authentication method (the password in the case of password authentication).
- The server authenticates the client. If the authentication fails, the server informs the client by sending a message, which includes a list of available methods for re-authentication.
- The client selects a method from the list to initiate another authentication.
- The above process repeats until the authentication succeeds or the authentication times timeout and the session is torn down.

SSH provides two authentication methods: password authentication and publickey authentication.

In password authentication:

- The client encrypts the username and password, encapsulates them into a password authentication request, and sends the request to the server.
- Upon receiving the request, the server decrypts the username and password, compares them against those it maintains, and then informs the client of the authentication result.

In publickey authentication:

- The server authenticates clients using digital signatures. Currently, the device supports two publickey algorithms to implement digital signatures: RSA and DSA. The client sends to the server a public authentication request containing its user name, public key and algorithm. The server validates the public key. If the public key is invalid, the authentication fails; otherwise, the server generates a digital signature to authenticate the client, and then sends back a message to inform the success or failure of the authentication.



Besides password authentication and publickey authentication, SSH provides another two authentication methods

- **password-publickey:** Performs both password authentication and publickey authentication of the client. A client running SSH1 client only needs to pass either type of the two, while a client running SSH2 client must pass both of them to login.
- **any:** Performs either password authentication or publickey authentication. The client tries publickey authentication first.

Session request

After passing authentication, the client sends a session request to the server, while the server listens to and processes the request from the client. If the client passes authentication, the server sends back to the client an SSH_MSG_SUCCESS packet and goes on to the interactive session stage with the client. Otherwise, the server sends back to the client an SSH_MSG_FAILURE packet, indicating that the processing fails or it cannot resolve the request.

Interactive session

In this stage, the server and the client exchanges data in this way:

- The client encrypts and sends the command to be executed to the server.
- The server decrypts and executes the command, and then encrypts and sends the result to the client.
- The client decrypts and displays the result on the terminal.



- *During interactive session, the client can send the commands to be performed by pasting the text, which must be within 2000 bytes. It is recommended that the text pasted be commands in the same view; otherwise, the server may not be able to perform the commands.*
- *If the text exceeds 2000 bytes, you can upload the configuration file to the server and use the configuration file to restart the server so that the server executes the commands.*

Configuring the Device as an SSH Server

SSH Server Configuration Task List

Complete the following tasks to configure an SSH server:

Task	Remarks	
"Enabling SSH Server" on page 1111	Required	
"Configuring the User Interfaces for SSH Clients" on page 1111	Required	
"Configuring RSA and DSA Keys" on page 1111	"Creating RSA or DSA key pairs" on page 1111	Required
	"Exporting RSA or DSA key pairs" on page 1112	Optional
	"Destroying RSA or DSA key pairs" on page 1112	Optional
"Configuring a Client Public Key" on page 1112	Required for publickey authentication users and optional for password authentication users	
"Configuring an SSH User" on page 1113	Optional	

Task	Remarks
"Setting the SSH Management Parameters" on page 1115	Optional



As a client uses either RSA or DSA algorithm for authentication and different clients may support different algorithms, the server needs to generate both RSA and DSA key pairs for successful authentication.

Enabling SSH Server

Follow these steps to enable SSH server:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the SSH server function	ssh server enable	Required Disabled by default

Configuring the User Interfaces for SSH Clients

An SSH client accesses the device through a VTY user interface. Therefore, you need to configure the user interfaces for SSH clients to allow SSH login. Note that the configuration takes effect at the next login.

Follow these steps to configure the protocols for the current user interface to support:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter user interface view of one or more user interfaces	user-interface vty <i>number</i> [<i>ending-number</i>]	Required
Set the login authentication method to scheme	authentication-mode scheme [command-authorization]	Required By default, the authentication mode is password .
Specify the protocols for the user interfaces to support	protocol inbound { all ssh telnet }	Optional All protocols are supported by default.



CAUTION:

- For detailed information about the **authentication-mode** and **protocol inbound** commands, refer to "Logging In to an Ethernet Switch" on page 27.
- If you configure a user interface to support SSH, be sure to configure the corresponding authentication method with the **authentication-mode scheme** command.
- For a user interface configured to support SSH, you cannot configure the **authentication-mode password** command and the **authentication-mode none** command.

Configuring RSA and DSA Keys

Creating RSA or DSA key pairs

For successful SSH login, you must create the RSA or DSA key pairs first.

Follow these steps to create an RSA or DSA key pair:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create the local RSA key pair	public-key local create rsa	Required
Create the local DSA key pair	public-key local create dsa	Use either command. By default, there is neither RSA key pair nor DSA key pair.

**CAUTION:**

- Configuration of the **rsa local-key-pair create** and **public-key local create dsa** command can survive a reboot. You only need to configure it once.
- The length of an RSA server/host key is in the range 512 to 2048 bits. With SSH2, however, some clients require that the keys generated by the server must not be less than 768 bits.
- The length of a DSA host key is in the range 512 to 2048 bits. With SSH2, nevertheless, some clients require that the keys generated by the server must not be less than 768 bits.

Exporting RSA or DSA key pairs

You can display or export the local RSA or DSA host key for setting the host key on the remote end.

Follow these steps to display or export an RSA or DSA host key:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Display the local RSA host key on the screen in a specified format, or export it to a specified file	public-key local export rsa { openssh ssh1 ssh2 } [<i>filename</i>]	Required Use either command.
Display the local DSA host key on the screen in a specified format, or export it to a specified file	public-key local export dsa { openssh ssh2 } [<i>filename</i>]	

Destroying RSA or DSA key pairs

Follow these steps to destroy an RSA or DSA key pair:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Destroy the local RSA key pair	public-key local destroy rsa	Required
Destroy the local DSA key pair	public-key local destroy dsa	Use either command.

Configuring a Client Public Key

This configuration task is only necessary for SSH users using publickey authentication.

For an SSH user that uses publickey authentication to login, the server must be configured with the client RSA or DSA host public key in advance, and the corresponding private key for the client must be specified on the client.

You can manually configure or import the publickey public key from a public key file. In the former case, you can manually copy the client's public key configuration to the server. In the latter case, the system automatically converts the public key to a string coded using the PKCS standard. Before importing the public key, you must upload the public key file (in binary) to the server through FTP or TFTP.



CAUTION:

- *When the device functions as the SSH server, you cannot use Secure CRT 4.07 to upload the client public key to the server.*
- *You can configure at most 20 client public keys on an SSH server.*

Configuring a client public key manually

Follow these steps to configure the client public key manually:

To do...	Use the command...	Remarks
Enter system view	System-view	-
Enter public key view	public-key peer <i>keyname</i>	-
Enter public key code view	public-key-code begin	-
Configure a client public key	Enter the content of the public key	Required The content must be a hexadecimal string that is generated randomly by the SSH-supported client software and coded compliant to PKCS. Spaces and carriage returns are allowed between characters.
Return from public key code view to public key view	public-key-code end	- When you exit public key code view, the system automatically saves the public key.
Return from public key view to system view	peer-public-key end	-

Importing a client public key from a public key file

Follow these steps to import a public key from a public key file:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Import the public key from a public key file	public-key peer <i>keyname</i> import sshkey <i>filename</i>	Required

Configuring an SSH User

This configuration allows you to create an SSH user and specify the service type and authentication method.

Follow these steps to configure an SSH user:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Create an SSH user, and specify the service type and authentication method	For stelnet users	ssh user <i>username</i> service-type stelnet authentication-type { password { any password-publickey publickey } assign publickey <i>keyname</i> }	Required Use either command.
	For all users or sftp users	ssh user <i>username</i> service-type { all sftp } authentication-type { password { any password-publickey publickey } assign publickey <i>keyname</i> work-directory <i>directory-name</i> }	

**CAUTION:**

- After passing AAA authentication, an AAA user without SSH user account still can log on to the server using password authentication and Stelnet or SFTP service.
- An SSH server supports up to 1024 SSH users.
- The service type of an SSH user can be Stelnet or SFTP. **stelnet**, or the secure Telnet protocol, refers to the traditional SSH service. For information about **stelnet**, refer to “SSH2.0 Overview” on page 1107. **sftp** represents the secure FTP protocol. For information about **sftp**, refer to “SFTP Overview” on page 1131.
- For successful login through SFTP, you must set the user service type to **sftp** or **all**.
- You can set the service type of an SSH user to **stelnet** or **all** if the user does not need SFTP service.
- As SSH1 does not support service type **sftp**, if the client uses SSH1 to log in to the server, you must set the service type to **stelnet** or **all** on the server. Otherwise, the client will fail to log in successfully.
- The working folder of an SFTP user is subject to the user authentication method. For a user using only password authentication, the working folder is the AAA authorized one. For a user using only publickey authentication or using both the publickey and password authentication methods, the working folder is the one set by using the **ssh user** command.
- The configured authentication method takes effect when the user logs in next time.



For users using publickey authentication:

- You must configure on the device the corresponding username and public keys.

- After login, the commands available for a user are determined by the user privilege level, which is configured with the **user privilege level** command on the user interface. By default, the command privilege level is 0.

For users using password authentication:

- You can configure the accounting information either on the device or on the remote authentication server (such as RADIUS authentication server).
- After login, the commands available to a user are determined by AAA authorization.

Setting the SSH Management Parameters

SSH management includes:

- Enabling the SSH server to be compatible with SSH1
- Setting the server key pair update interval, applicable to users using SSH1 client.
- Setting the SSH user authentication timeout period
- Setting the maximum number of SSH authentication attempts

Setting the above parameters can help avoid malicious guess at and cracking of the keys and usernames, securing your SSH connections.

Follow these steps to set the SSH management parameters:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the SSH server to work with SSH1.x clients	ssh server compatible-ssh1x enable	Optional By default, the SSH server can work with SSH1.x clients.
Set the RSA server key pair update interval	ssh server rekey-interval <i>hours</i>	Optional 0 by default, that is, the RSA server key pair is not updated.
Set the SSH user authentication timeout period	ssh server authentication-timeout <i>time-out-value</i>	Optional 60 seconds by default
Set the maximum number of SSH authentication attempts	ssh server authentication-retries <i>times</i>	Optional 3 by default



*Authentication will fail if the number of authentication attempts (including both publickey and password authentication) exceeds that specified in the **ssh server authentication-retries** command.*

Configuring the Device as an SSH Client

SSH Client Configuration Task List

Complete the following tasks to configure an SSH client:

Task	Remarks
"Specifying a Source IP address/Interface for the SSH client" on page 1116	Optional
"Configuring Whether First-time Authentication is Supported" on page 1116	Optional
"Establishing a Connection Between the SSH Client and the Server" on page 1117	Required

Specifying a Source IP address/Interface for the SSH client

This configuration task allows you to specify a source IP address or interface for the client to access the SSH server, improving service manageability.

To do...	Use the command...	Remarks
Enter system view	system-view	-
Specify a source IP address or interface for the SSH client	ssh client source { ip <i>ip-address</i> interface <i>interface-type</i> <i>interface-number</i> }	Required By default, the address of the interface decided by the routing is used to access the SSH server
Specify a source IPv6 address or interface for the SSH client	ssh client ipv6 source { ipv6 <i>ipv6-address</i> interface <i>interface-type</i> <i>interface-number</i> }	

Configuring Whether First-time Authentication is Supported

When the device connects to the SSH server as an SSH client, you can configure whether the device supports first-time authentication.

- With first-time authentication, when an SSH client not configured with the server host public key accesses the server for the first time, the user can continue accessing the server, and save the host public key on the client for use in subsequent authentications.
- Without first-time authentication, a client not configured with the server host public key will be denied of access to the server. To access the server, a user must configure in advance the server host public key locally and specify the public key name for authentication.

Enable the device to support first-time authentication

Follow these steps to enable the device to support first-time authentication:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the device to support first-time authentication	ssh client first-time enable	Optional By default, first-time authentication is supported on a client.

Disable first-time authentication

For successful authentication of an SSH client not supporting first-time authentication, the server host public key must be configured on the client and the public key name must be specified.

Follow these steps to disable first-time authentication:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Disable first-time authentication support	undo ssh client first-time	Optional By default, first-time authentication is supported on a client.
Configure the server public key	Refer to "Configuring a Client Public Key" on page 1112 "Configuring a Client Public Key" on page 1112	Required The method of configuring server public key on the client is similar to that of configuring client public key on the server.
Specify the host public key name of the server	ssh client authentication server server assign publickey keyname	Required

Establishing a Connection Between the SSH Client and the Server

Follow these steps to establish the connection between the SSH client and the server:

To do...		Use the command...	Remarks
Establish a connection between the SSH client and the server, and specify the preferred key exchange algorithm, encryption algorithms, and HMAC algorithms for them	Establish a connection between the SSH client and the IPv4 server, and specify the preferred key exchange algorithm, encryption algorithms, and HMAC algorithms for them	ssh2 server [<i>port-number</i>] [identity-key { dsa rsa }] prefer-ctos-cipher { aes128 des } prefer-ctos-hmac { md5 md5-96 sha1 sha1-96 } prefer-kex { dh-group-exchange dh-group1 dh-group14 } prefer-stoc-cipher { aes128 des } prefer-stoc-hmac { md5 md5-96 sha1 sha1-96 }] *	Required Use either command in user view.
	Establish a connection between the SSH client and the IPv6 server, and specify the preferred key exchange algorithm, encryption algorithms, and HMAC algorithms for them	ssh2 ipv6 server [<i>port-number</i>] [identity-key { dsa rsa }] prefer-ctos-cipher { aes128 des } prefer-ctos-hmac { md5 md5-96 sha1 sha1-96 } prefer-kex { dh-group-exchange dh-group1 dh-group14 } prefer-stoc-cipher { aes128 des } prefer-stoc-hmac { md5 md5-96 sha1 sha1-96 }] *	

Displaying and Maintaining SSH

To do...	Use the command...	Remarks
Display information about the public keys of the local key pair	display public-key local { dsa rsa } public	Available in any view
Display information about the public keys	display public-key peer [brief name <i>publickey-name</i>]	Available in any view
Display the source IP address or interface currently set for the SFTP client	display sftp client source	Available in any view
Display the source IP address or interface currently set for the SSH client	display ssh client source	Available in any view
Display the status information or session information of an SSH server	display ssh server { status session }	Available in any view

To do...	Use the command...	Remarks
Display the mappings between host public keys and SSH servers saved on a client	display ssh server-info	Available in any view
Display information about a specified or all SSH users on the SSH server	display ssh user-information [<i>username</i>]	Available in any view

SSH Server Configuration Examples

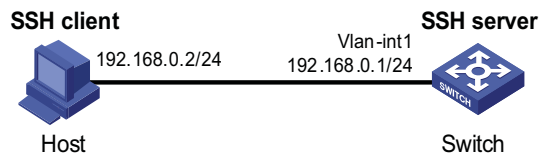
When Using Password Authentication

Network requirements

- As shown in Figure 331, a local SSH connection is established between the host (SSH client) and the switch (SSH server) for secure data exchange.
- Password authentication is required.

Network diagram

Figure 331 Network diagram for SSH server configuration (using password authentication)



Configuration procedure

Configure the SSH server

Generate RSA and DSA key pairs and enable the SSH server.

```

<Switch> system-view
[Switch] public-key local create rsa
[Switch] public-key local create dsa
[Switch] ssh server enable
  
```

Configure an IP address for VLAN interface 1. This address will serve as the destination for the SSH client in connecting the server.

```

[Switch] interface vlan-interface 1
[Switch-Vlan-interface1] ip address 192.168.1.40 255.255.255.0
[Switch-Vlan-interface1] quit
  
```

Set the authentication mode for the user interface to AAA.

```

[Switch] user-interface vty 0 4
[Switch-ui-vty0-4] authentication-mode scheme
  
```

Enable the user interface to support SSH.

```

[Switch-ui-vty0-4] protocol inbound ssh
[Switch-ui-vty0-4] quit
  
```

Create local user client001, and set the user command privilege level to 3

```
[Switch] local-user client001
[Switch-luser-client001] password simple aabbcc
[Switch-luser-client001] service-type ssh level 3
[Switch-luser-client001] quit
```

Specify the service type for user client001 as Stelnet, and the authentication method as password.

```
[Switch] ssh user client001 service-type stelnet authentication-type password
```

Configure the SSH client

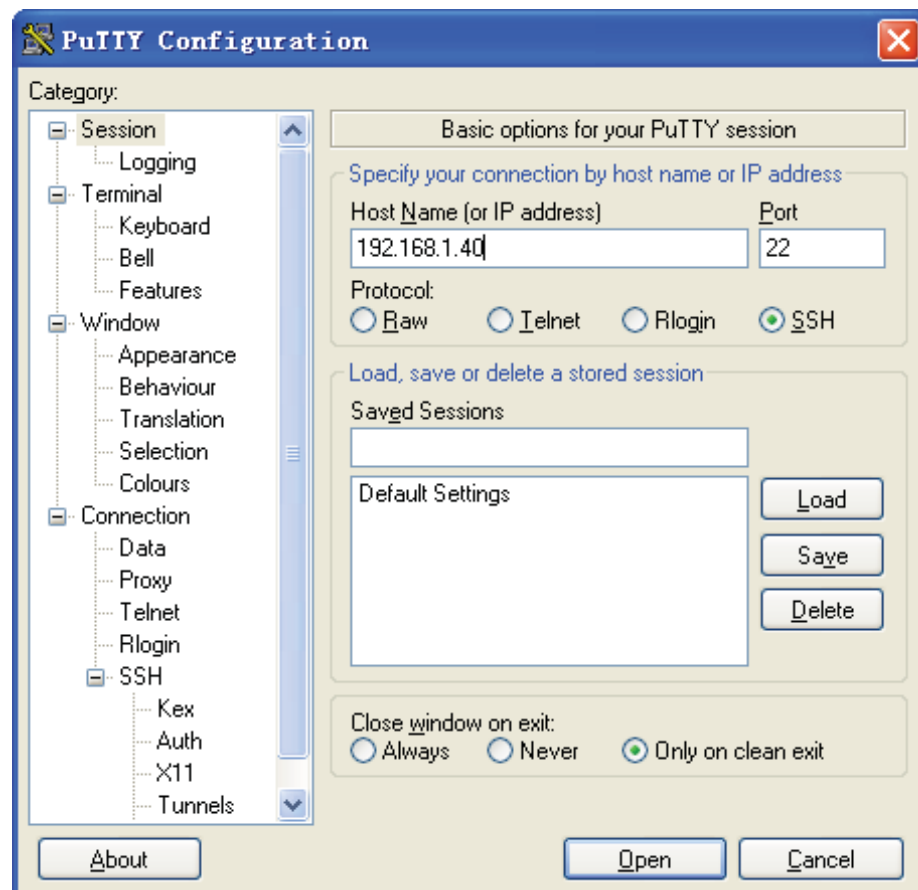


There are a variety of SSH client software, such as PuTTY, OpenSSH, and so on. The following is an example of configuring SSH client using PuTT v0.58.

Establish a connection with the SSH server

Launch PuTTY.exe to enter the following interface. In the **Host Name (or IP address)** text box, enter the IP address of the server (192.168.1.40).

Figure 332 SSH client configuration interface



From the window shown in Figure 332, click **Open**. The following SSH client interface appears. If the connection is normal, you will be prompted to enter the username (client001) and password (aabbcc)

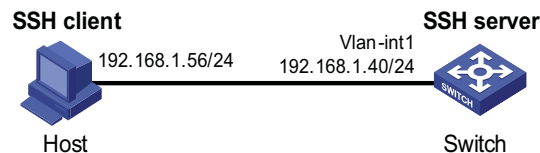
When Using Publickey Authentication

Network requirements

- As shown in Figure 333, a local SSH connection is established between the host (SSH client) and the switch (SSH server) for secure data exchange.
- Publickey authentication is used, the algorithm is RSA.

Network diagram

Figure 333 Network diagram of SSH server configuration (using publickey authentication)



Configuration procedure

1 Configure the SSH server

Generate RSA and DSA key pairs and enable SSH server.

```
<Switch> system-view
[Switch] public-key local create rsa
[Switch] public-key local create dsa
[Switch] ssh server enable
```

Configure an IP address for VLAN interface 1. This address will serve as the destination for the SSH client in connecting the server.

```
[Switch] interface vlan-interface 1
[Switch-Vlan-interface1] ip address 192.168.1.40 255.255.255.0
[Switch-Vlan-interface1] quit
```

Set the authentication mode for the user interface to AAA.

```
[Switch] user-interface vty 0 4
[Switch-ui-vty0-4] authentication-mode scheme
```

Enable the user interface to support SSH.

```
[Switch-ui-vty0-4] protocol inbound ssh
```

Set the user command privilege level to 3.

```
[Switch-ui-vty0-4] user privilege level 3
[Switch-ui-vty0-4] quit
```



Before performing the following tasks, you must generate an RSA public key pair (using the client software) on the client, save the key pair in a file named key.pub, and then upload the file to the SSH server through FTP or TFTP. For details, refer to "Configuring the Device as an SSH Client" on page 1115.

Import the client's public key from file "key.pub".

```
[Switch] public-key peer Switch001 import sshkey key.pub
```

Specify the authentication type for user "client002" as publickey, and assign the public key "Switch001" for the user.

```
[Switch] ssh user client002 service-type stelnet authentication-type
publickey assign publickey Switch001
```

2 Configure the SSH client

Generate an RSA key pair

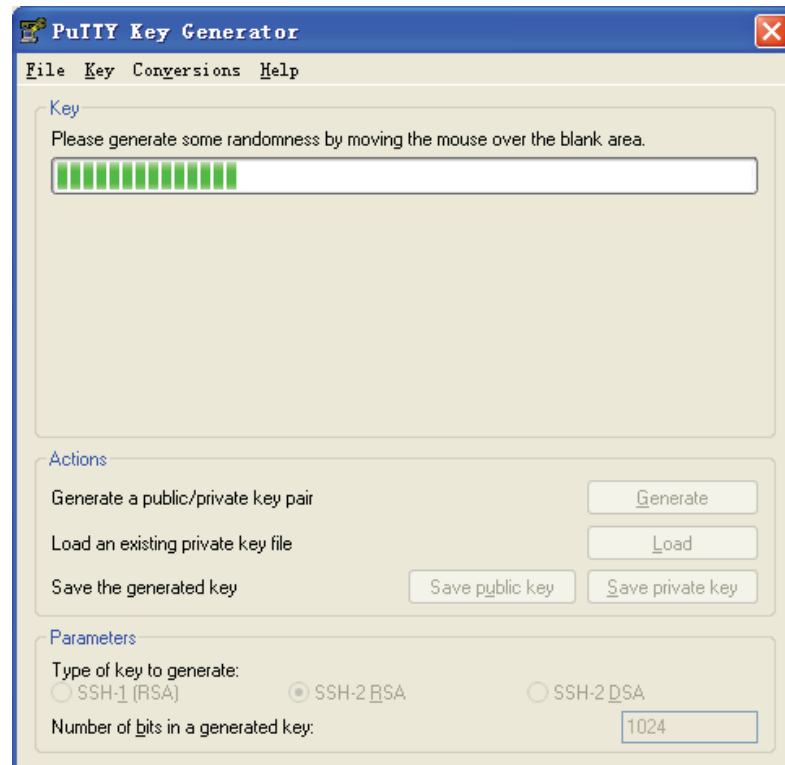
Run PuTTYGen.exe, choose **SSH2-(RSA)** and click **Generate**.

Figure 334 Generate a client key pair (1)



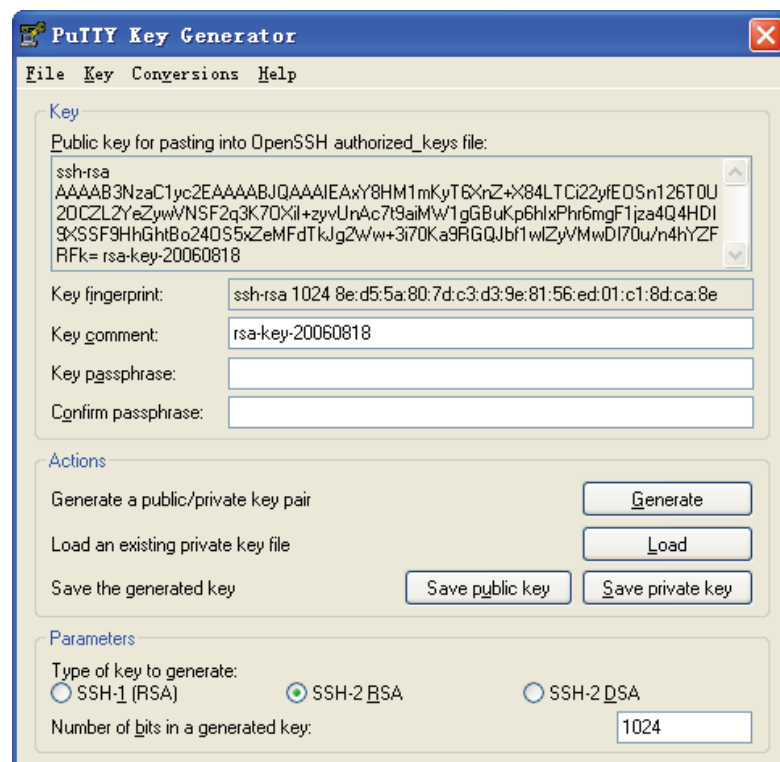
While generating the key pair, you must move the mouse continuously and keep the mouse off the green process bar shown in Figure 335. Otherwise, the process bar stops moving and the key pair generating process is stopped.

Figure 335 Generate a client key pair (2)



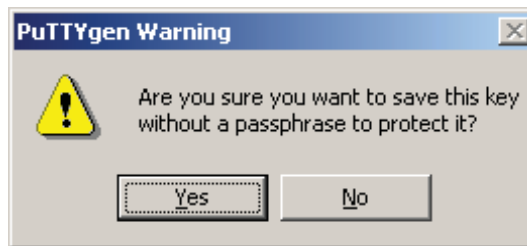
After the key pair is generated, click **Save public key** to save the key in a file by entering a file name ("key.pub" in this case).

Figure 336 Generate a client key pair (3)



Likewise, to save the private key, click **Save private key**. A warning window pops up to prompt you whether to save the private key without any protection. Click **Yes** and enter the name of the file for saving the key (“private” in this case).

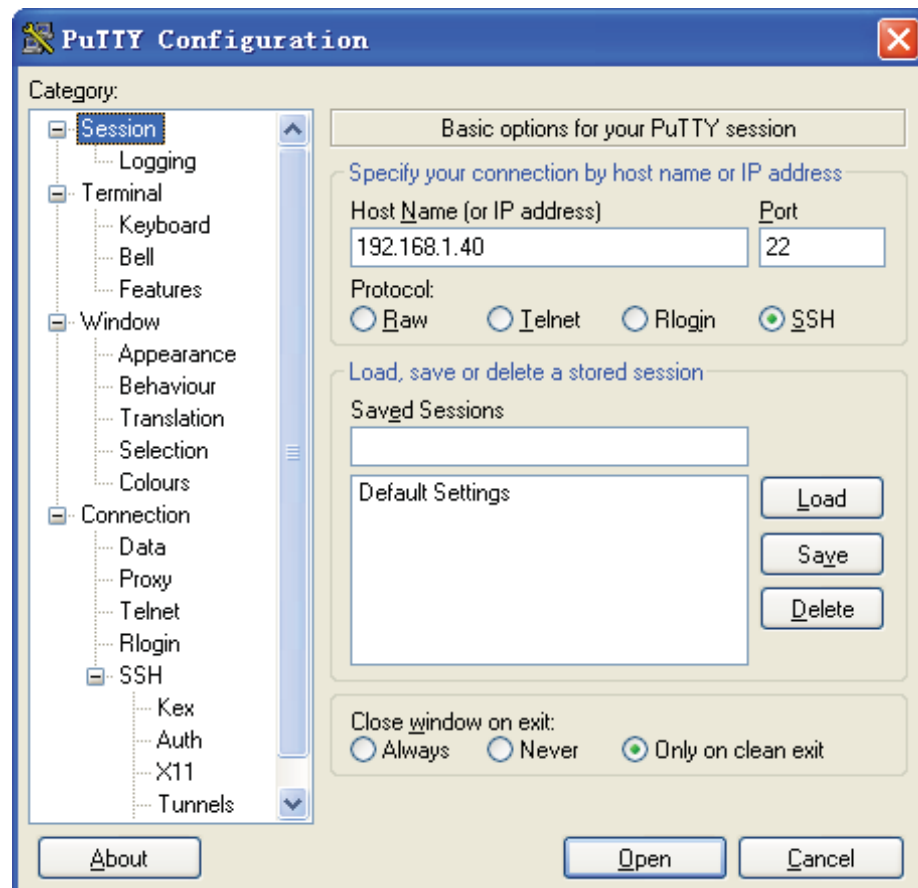
Figure 337 Generate a client key pair (4)



After generating a key pair on a client, you need to transmit the saved public key file to the server through FTP or TFTP and have the configuration on the server done before continuing configuration of the client.

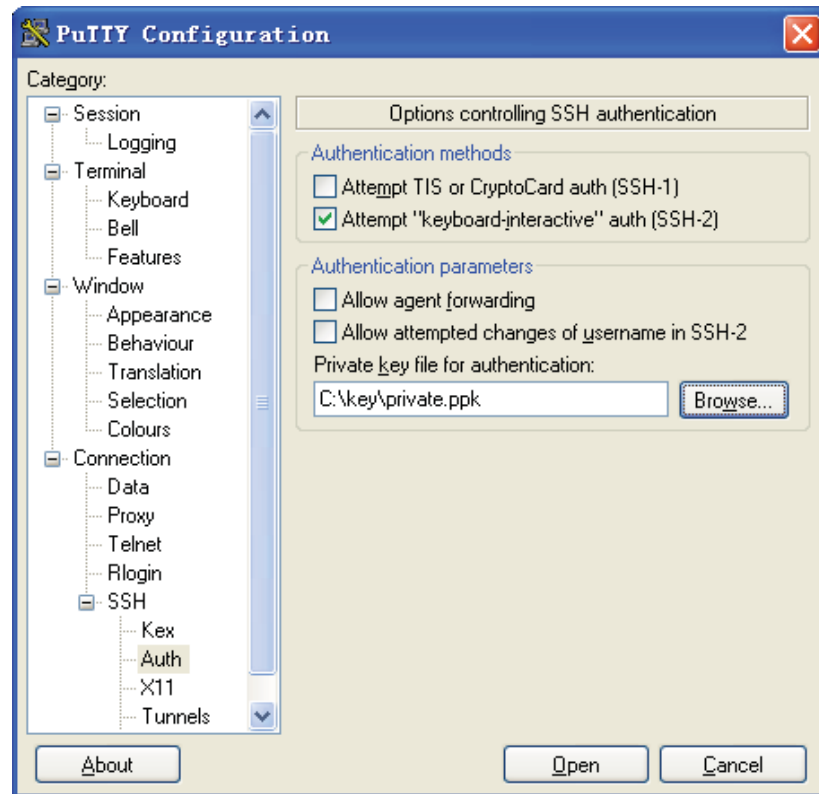
Specify the private key file and establish a connection with the SSH server
Launch PuTTY.exe to enter the following interface. In the **Host Name (or IP address)** text box, enter the IP address of the server (192.168.1.40).

Figure 338 SSH client configuration interface (1)



Select **Connection/SSH/Auth**. The following window appears. Click **Browse...** to bring up the file selection window, navigate to the private key file and click **OK**.

Figure 339 SSH client configuration interface (2)



From the window shown in Figure 339, click **Open**. The following SSH client interface appears. If the connection is normal, you will be prompted to enter the username (client002) to enter the configuration interface.

SSH Client Configuration Examples

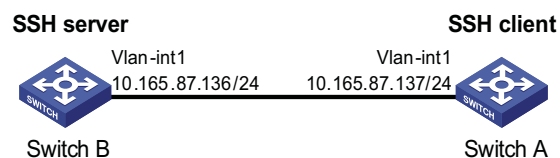
When Using Password Authentication

Network requirements

- As shown in Figure 340, Switch A (the SSH client) needs to log on to Switch B (the SSH server) through the SSH protocol.
- The username of the SSH client is **client001** and the password is **aabbcc**. Password authentication is required.

Network diagram

Figure 340 Network diagram for SSH client configuration (using password authentication)



Configuration procedure**1** Configure the SSH server

Create an RSA and DSA key pair and enable the SSH server.

```
<SwitchB> system-view
[SwitchB] public-key local create rsa
[SwitchB] public-key local create dsa
[SwitchB] ssh server enable
```

Create an IP address for VLAN interface 1, which the SSH client will use as the destination for SSH connection.

```
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ip address 10.165.87.136 255.255.255.0
[SwitchB-Vlan-interface1] quit
```

Set the authentication mode for the user interface to AAA.

```
[SwitchB] user-interface vty 0 4
[SwitchB-ui-vty0-4] authentication-mode scheme
```

Enable the user interface to support SSH.

```
[SwitchB-ui-vty0-4] protocol inbound ssh
[SwitchB-ui-vty0-4] quit
```

Create local user client001.

```
[SwitchB] local-user client001
[SwitchB-luser-client001] password simple aabbcc
[SwitchB-luser-client001] service-type ssh level 3
[SwitchB-luser-client001] quit
```

Specify the service type for user "client001" as Stelnet, and the authentication method as password.

```
[SwitchB] ssh user client001 service-type stelnet authentication-type password
```

2 Configure the SSH client

Configure an IP address for VLAN interface 1.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ip address 10.165.87.137 255.255.255.0
[SwitchA-Vlan-interface1] quit
```

Disable first-time authentication.

```
[SwitchA] undo ssh client first-time
```

Configure the host public key of the SSH server.

```
[SwitchA] public-key peer key1
[SwitchA-pkey-public-key] public-key-code begin
[SwitchA-pkey-key-code]308201B73082012C06072A8648CE3804013082011F0281810
0D757262C4584C44C211F18BD96E5F0
[SwitchA-pkey-key-code]61C4F0A423F7FE6B6B85B34CEF72CE14A0D3A5222FE08CECE
65BE6C265854889DC1EDBD13EC8B274
[SwitchA-pkey-key-code]DA9F75BA26CCB987723602787E922BA84421F22C3C89CB9B0
6FD60FE01941DDD77FE6B12893DA76E
[SwitchA-pkey-key-code]EBC1D128D97F0678D7722B5341C8506F358214B16A2FAC4B3
68950387811C7DA33021500C773218C
[SwitchA-pkey-key-code]737EC8EE993B4F2DED30F48EDACE915F0281810082269009E
14EC474BAF2932E69D3B1F18517AD95
[SwitchA-pkey-key-code]94184CCDFCEAE96EC4D5EF93133E84B47093C52B20CD35D02
492B3959EC6499625BC4FA5082E22C5
[SwitchA-pkey-key-code]B374E16DD00132CE71B020217091AC717B612391C76C1FB2E
```

```

88317C1BD8171D41ECB83E210C03CC9
[SwitchA-pkey-key-code]B32E810561C21621C73D6DAAC028F4B1585DA7F42519718CC
9B09EEF0381840002818000AF995917
[SwitchA-pkey-key-code]E1E570A3F6B1C2411948B3B4FFA256699B3BF871221CC9C5D
F257523777D033BEE77FC378145F2AD
[SwitchA-pkey-key-code]D716D7DB9FCABB4ADB6FB4FDB0CA25C761B308EF53009F71
01F7C62621216D5A572C379A32AC290
[SwitchA-pkey-key-code]E55B394A217DA38B65B77F0185C8DB8095522D1EF044B465E
8716261214A5A3B493E866991113B2D
[SwitchA-pkey-key-code]485348
[SwitchA-pkey-key-code] public-key-code end
[SwitchA-pkey-public-key] peer-public-key end

# Specify the host public key for the SSH server (10.165.87.136) as "key1".

[SwitchA] ssh client authentication server 10.165.87.136 assign publ
ickey key1
[SwitchA] quit

# Establish an SSH connection to server 10.165.87.136.

<SwitchA> ssh2 10.165.87.136
Username: client001
Trying 10.165.87.136
Press CTRL+K to abort
Connected to 10.165.87.136...
Enter password:
*****
* Copyright (c) 2004-2008 3Com Corporation. All rights reserved. *
* Without the owner's prior written consent, *
* no decompiling or reverse-switch fabricering shall be allowed. *
*****

<SwitchB>

```

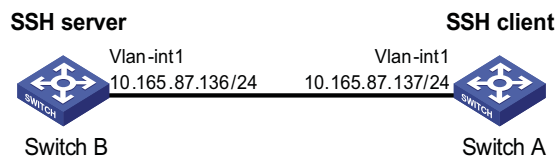
When Using Publickey Authentication

Network requirements

- As shown in Figure 341, Switch A (the SSH client) needs to log on to Switch B (the SSH server) through SSH protocol.
- Publickey authentication is used; the algorithm is DSA.

Network diagram

Figure 341 Network diagram of SSH client configuration (using publickey authentication)



Configuration procedure

1 Configure the SSH server

Generate RSA and DSA key pairs and enable SSH server.

```

<SwitchB> system-view
[SwitchB] public-key local create rsa
[SwitchB] public-key local create dsa
[SwitchB] ssh server enable

```

Configure an IP address for VLAN interface 1, which the SSH client will use as the destination for SSH connection.

```
[SwitchB] interface vlan-interface 1
[SwitchB-Vlan-interface1] ip address 10.165.87.136 255.255.255.0
[SwitchB-Vlan-interface1] quit
```

Set the authentication mode for the user interface to AAA.

```
[SwitchB] user-interface vty 0 4
[SwitchB-ui-vty0-4] authentication-mode scheme
```

Enable the user interface to support SSH.

```
[SwitchB-ui-vty0-4] protocol inbound ssh
```

Set the user command privilege level to 3.

```
[SwitchB-ui-vty0-4] user privilege level 3
[SwitchB-ui-vty0-4] quit
```



Before performing the following tasks, you must generate a DSA public key pair (using the client software) on the client, save the key pair in a file named key.pub, and then upload the file to the SSH server through FTP or TFTP. For details, refer to “Configuring the Device as an SSH Client” on page 1115.

Import the remote public key pair from the file “key.pub”.

```
[SwitchB] public-key peer Switch001 import sshkey key.pub
```

Specify the authentication type for user “client002” as publickey, and assign the public key “Switch001” for the user.

```
[SwitchB] ssh user client002 service-type stelnet authentication-type publickey assign publickey Switch001
```

2 Configure the SSH client

Configure an IP address for Vlan interface 1.

```
<SwitchA> system-view
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ip address 10.165.87.137 255.255.255.0
[SwitchA-Vlan-interface1] quit
```

Generate a DSA key pair.

```
[SwitchA] public-key local create dsa
```

Export the DSA key pair to the file **key.pub**.

```
[SwitchA] public-key local export dsa ssh2 key.pub
[SwitchA] quit
```



After generating a key pair on a client, you need to transmit the saved public key file to the server through FTP or TFTP and have the configuration on the server done before continuing configuration of the client.

Establish an SSH connection to the server (10.165.87.136).

```
<SwitchA> ssh2 10.165.87.136
Username: client002
Trying 10.165.87.136 ...
Press CTRL+K to abort
Connected to 10.165.87.136 ...
```

```
The Server is not authenticated. Continue? [Y/N]:y
Do you want to save the server public key? [Y/N]:n
```

```
*****
* Copyright(c) 2004-2008 3Com Corporation. All rights reserved. *
* Without the owner's prior written consent, *
* no decompiling or reverse-switch fabricering shall be allowed. *
*****
```

<SwitchB>

When configuring SFTP, go to these sections for information you are interested in:

- “SFTP Overview” on page 1131
- “Configuring an SFTP Server” on page 1131
- “Configuring an SFTP Client” on page 1132
- “SFTP Configuration Example” on page 1135

SFTP Overview

The secure file transfer protocol (SFTP) is a new feature in SSH 2.0.

SFTP uses the SSH connection to provide secure data transfer. The device can serve as the SFTP server, allowing a remote user to login to the SFTP server for secure file management and transfer. The device can also server as an SFTP client, enabling a user to login from the device to a remote device for secure file transfer.

Configuring an SFTP Server

Configuration Prerequisites

- You have configured the SSH server. For the detailed configuration procedure, refer to “Configuring the Device as an SSH Server” on page 1110.
- You have used the **ssh user service-type** command to set the service type of SSH users to **sftp** or **all**. For configuration procedure, refer to “Configuring an SSH User” on page 1113.

Enabling the SFTP Server

This configuration task is to enable the SFTP service so that a client can login to the SFTP server through SFTP.

Follow these steps to enable the SFTP server:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the SFTP server	sftp server enable	Required Disabled by default



When the device functions as the SFTP server, only one client can access the SFTP server at a time. If the SFTP client uses WinSCP, a file on the server cannot be modified directly; it can only be downloaded to a local place, modified, and then uploaded to the server.

Configuring the SFTP Connection Idle Timeout Period

Once the idle period of an SFTP connection exceeds the specified threshold, the system automatically tears the connection down, so that a user cannot occupy a connection for nothing.

Follow these steps to configure the SFTP connection idle timeout period:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the SFTP connection idle timeout period	sftp server idle-timeout <i>time-out-value</i>	Required 10 minutes by default

Configuring an SFTP Client

Specifying a Source IP Address or Interface for the SFTP Client

You can configure a client to use only a specified source IP address or interface to access the SFTP server, thus enhancing the service manageability.

Follow these steps to specify a source IP address or interface for the SFTP client:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Specify a source IP address or interface for the SFTP client	Specify a source IPv4 address or interface for the SFTP client sftp client source { ip <i>ip-address</i> interface <i>interface-type</i> <i>interface-number</i> }	Required Use either command. By default, an SFTP client uses the interface address specified by the route of the device to access the SFTP server.
Specify a source IPv6 address or interface for the SFTP client	Specify a source IPv6 address or interface for the SFTP client sftp client ipv6 source { <i>ipv6 ipv6-address</i> interface <i>interface-type</i> <i>interface-number</i> }	

Establishing a Connection to the SFTP Server

This configuration task is to enable the SFTP client to establish a connection with the remote SFTP server and enter SFTP client view.

Follow these steps to enable the SFTP client:

To do...		Use the command...	Remarks
Establish a connection to the remote SFTP server and enter SFTP client view	Establish a connection to the remote IPv4 SFTP server and enter SFTP client view	sftp server [<i>port-number</i>] [identity-key { dsa rsa }] prefer-ctos-cipher { aes128 des } prefer-ctos-hmac { md5 md5-96 sha1 sha1-96 } prefer-kex { dh-group-exchange dh-group1 dh-group14 } prefer-stoc-cipher { aes128 des } prefer-stoc-hmac { md5 md5-96 sha1 sha1-96 }] *	Required Use either command in user view.
	Establish a connection to the remote IPv6 SFTP server and enter SFTP client view	sftp ipv6 server [<i>port-number</i>] [identity-key { dsa rsa }] prefer-ctos-cipher { aes128 des } prefer-ctos-hmac { md5 md5-96 sha1 sha1-96 } prefer-kex { dh-group-exchange dh-group1 dh-group14 } prefer-stoc-cipher { aes128 des } prefer-stoc-hmac { md5 md5-96 sha1 sha1-96 }] *	

Working with the SFTP Directories

SFTP directory operations include:

- Changing or displaying the current working directory
- Displaying files under a specified directory or the directory information
- Changing the name of a specified directory on the server
- Creating or deleting a directory

Follow these steps to work with the SFTP directories:

To do...	Use the command...	Remarks
Establish a connection to the remote SFTP server and enter SFTP client view	sftp [ipv6] server [port-number] [identity-key { dsa rsa }] prefer-ctos-cipher { aes128 des } prefer-ctos-hmac { md5 md5-96 sha1 sha1-96 } prefer-kex { dh-group-exchange dh-group1 dh-group14 } prefer-stoc-cipher { aes128 des } prefer-stoc-hmac { md5 md5-96 sha1 sha1-96 }] *	Required Execute the command in user view.
Change the working directory of the remote SFTP server	cd [remote-path]	Optional

To do...	Use the command...	Remarks
Return to the upper-level directory	cdup	Optional
Display the current working directory of the remote SFTP server	pwd	Optional
Display files under a specified directory	dir [-a -l] [<i>remote-path</i>] ls [-a -l] [<i>remote-path</i>]	Optional The dir command functions as the ls command.
Change the name of a specified directory on the SFTP server	rename <i>oldname newname</i>	Optional
Create a new directory on the remote SFTP server	mkdir <i>remote-path</i>	Optional
Delete a directory from the SFTP server	rmdir <i>remote-path</i> &<1-10>	Optional

Working with SFTP Files

SFTP file operations include:

- Changing the name of a file
- Downloading a file
- Uploading a file
- Displaying a list of the files
- Deleting a file

Follow these steps to work with SFTP files:

To do...	Use the command...	Remarks
Establish a connection to the remote SFTP server and enter SFTP client view	sftp [<i>ipv6</i>] <i>server</i> [<i>port-number</i>] [identity-key { <i>dsa</i> <i>rsa</i> } prefer-ctos-cipher { <i>aes128</i> <i>des</i> } prefer-ctos-hmac { <i>md5</i> <i>md5-96</i> <i>sha1</i> <i>sha1-96</i> } prefer-kex { <i>dh-group-exchange</i> <i>dh-group1</i> <i>dh-group14</i> } prefer-stoc-cipher { <i>aes128</i> <i>des</i> } prefer-stoc-hmac { <i>md5</i> <i>md5-96</i> <i>sha1</i> <i>sha1-96</i> }] *	Required Execute the command in user view.
Change the name of a specified file on the SFTP server	rename <i>old-name new-name</i>	Optional
Download a file from the remote server and save it locally	get <i>remote-file</i> [<i>local-file</i>]	Optional
Upload a local file to the remote SFTP server	put <i>local-file</i> [<i>remote-file</i>]	Optional
Display the files under a specified directory	dir [-a -l] [<i>remote-path</i>] ls [-a -l] [<i>remote-path</i>]	Optional The dir command functions as the ls command.

To do...	Use the command...	Remarks
Delete a file from the SFTP server	delete <i>remote-file</i> &<1-10> remove <i>remote-file</i> &<1-10>	Optional The delete command functions as the remove command.

Displaying Help Information

This configuration task is to display a list of all commands or the help information of an SFTP client command, such as the command format and parameters.

Follow these steps to display a list of all commands or the help information of an SFTP client command:

To do...	Use the command...	Remarks
Establish a connection to the remote SFTP server and enter SFTP client view	sftp [ipv6] <i>server</i> [<i>port-number</i>] [identity-key { dsa rsa }] prefer-ctos-cipher { aes128 des } prefer-ctos-hmac { md5 md5-96 sha1 sha1-96 } prefer-kex { dh-group-exchange dh-group1 dh-group14 } prefer-stoc-cipher { aes128 des } prefer-stoc-hmac { md5 md5-96 sha1 sha1-96 }] *	Required Execute the command in user view.
Display a list of all commands or the help information of an SFTP client command	help [all <i>command-name</i>]	Required

Terminating the Connection to the Remote SFTP Server

Follow these steps to terminate the connection to the remote SFTP server:

To do...	Use the command...	Remarks
Establish a connection to the remote SFTP server and enter SFTP client view	sftp [ipv6] <i>server</i> [<i>port-number</i>] [identity-key { dsa rsa }] prefer-ctos-cipher { aes128 des } prefer-ctos-hmac { md5 md5-96 sha1 sha1-96 } prefer-kex { dh-group-exchange dh-group1 dh-group14 } prefer-stoc-cipher { aes128 des } prefer-stoc-hmac { md5 md5-96 sha1 sha1-96 }] *	Required Execute the command in user view.
Terminate the connection to the remote SFTP server and return to user view	bye exit quit	Required. Use any of the commands. These three commands function in the same way.

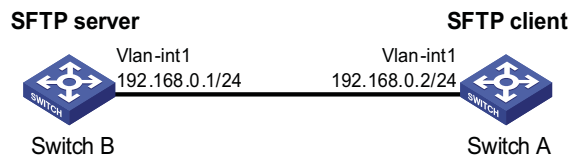
SFTP Configuration Example

Network requirements

As shown in Figure 342, an SSH connection is established between Switch A and Switch B. Switch A, an SFTP client, uses the username **client001** and password **aabbcc** to login to Switch B for file management and file transfer.

Network diagram

Figure 342 Network diagram for SFTP configuration



Configuration procedure

1 Configure the SFTP server (Switch B)

Generate RSA and DSA key pairs and enable the SSH server.

```

<SwitchB> system-view
[SwitchB] public-key local create rsa
[SwitchB] public-key local create dsa
[SwitchB] ssh server enable
  
```

Configure an IP address for VLAN interface 1, which the SSH client uses as the destination for SSH connection.

```

[SwitchB] interface Vlan-interface 1
[SwitchB-Vlan-interface1] ip address 192.168.0.1 255.255.255.0
[SwitchB-Vlan-interface1] quit
  
```

Set the authentication method on the user interface to AAA.

```

[SwitchB] user-interface vty 0 4
[SwitchB-ui-vty0-4] authentication-mode scheme
  
```

Set the protocol that a remote user uses to login as SSH.

```

[SwitchB-ui-vty0-4] protocol inbound ssh
[SwitchB-ui-vty0-4] quit
  
```

Create local user client001.

```

[SwitchB] local-user client001
[SwitchB-luser-client001] password simple aabbcc
[SwitchB-luser-client001] service-type ssh
[SwitchB-luser-client001] quit
  
```

Set the SSH authentication method to password, service type to SFTP.

```

[SwitchB] ssh user client001 service-type sftp authentication-type password
  
```



If you set the SSH authentication method to publickey, you need to configure the host public key of SwitchA. For the specific configuration, refer to “When Using Publickey Authentication” on page 1127.

Enable the SFTP server.

```

[SwitchB] sftp server enable
  
```

2 Configure the SFTP client (Switch A)

Configure an IP address for VLAN interface 1.

```

<SwitchA> system-view
[SwitchA] interface vlan-interface 1
[SwitchA-Vlan-interface1] ip address 192.168.0.2 255.255.255.0
[SwitchA-Vlan-interface1] quit
[SwitchA] quit
  
```

Establish a connection to the remote SFTP server and enter SFTP client view.

```
<SwitchA> sftp 192.168.0.1
Input Username: client001
Trying 192.168.0.1 ...
Press CTRL+K to abort
Connected to 192.168.0.1 ...
```

```
The Server is not authenticated. Continue? [Y/N]:y
Do you want to save the server public key? [Y/N]:y
Enter password:
```

```
sftp-client>
```

Display files under the current directory of the server, delete the file named "z", and check if the file is deleted successfully.

```
sftp-client> dir
-rwxrwxrwx  1 noone  nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx  1 noone  nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx  1 noone  nogroup       283 Aug 24 07:39 pubkey1
drwxrwxrwx  1 noone  nogroup         0 Sep 01 06:22 new
-rwxrwxrwx  1 noone  nogroup       225 Sep 01 06:55 pub
-rwxrwxrwx  1 noone  nogroup         0 Sep 01 08:00 z
```

```
sftp-client> delete z
The following files will be deleted:
```

```
/z
```

```
Are you sure to delete it? [Y/N]:y
```

```
This operation may take a long time.Please wait...
```

```
File successfully Removed
```

```
sftp-client> dir
-rwxrwxrwx  1 noone  nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx  1 noone  nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx  1 noone  nogroup       283 Aug 24 07:39 pubkey1
drwxrwxrwx  1 noone  nogroup         0 Sep 01 06:22 new
-rwxrwxrwx  1 noone  nogroup       225 Sep 01 06:55 pub
```

Add a directory named "new1" and check if it is created successfully.

```
sftp-client> mkdir new1
```

```
New directory created
```

```
sftp-client> dir
-rwxrwxrwx  1 noone  nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx  1 noone  nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx  1 noone  nogroup       283 Aug 24 07:39 pubkey1
drwxrwxrwx  1 noone  nogroup         0 Sep 01 06:22 new
-rwxrwxrwx  1 noone  nogroup       225 Sep 01 06:55 pub
drwxrwxrwx  1 noone  nogroup         0 Sep 02 06:30 new1
```

Rename directory "new1" to "new2" and check if the directory is renamed successfully.

```
sftp-client> rename new1 new2
```

```
File successfully renamed
```

```
sftp-client> dir
-rwxrwxrwx  1 noone  nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx  1 noone  nogroup       225 Aug 24 08:01 pubkey2
-rwxrwxrwx  1 noone  nogroup       283 Aug 24 07:39 pubkey1
drwxrwxrwx  1 noone  nogroup         0 Sep 01 06:22 new
```

```
-rwxrwxrwx  1 noone  nogroup      225 Sep 01 06:55 pub
drwxrwxrwx  1 noone  nogroup      0 Sep 02 06:33 new2
```

Download the file "pubkey2" from the server and change the name to "public".

```
sftp-client> get pubkey2 public
Remote file:/pubkey2 ---> Local file: public
Downloading file successfully ended
```

Upload the local file "pu" to the server, save it as "puk", and check if the file is uploaded successfully.

```
sftp-client> put pu puk
Local file:pu ---> Remote file: /puk
Uploading file successfully ended
sftp-client> dir
-rwxrwxrwx  1 noone  nogroup      1759 Aug 23 06:52 config.cfg
-rwxrwxrwx  1 noone  nogroup      225 Aug 24 08:01 pubkey2
-rwxrwxrwx  1 noone  nogroup      283 Aug 24 07:39 pubkey1
drwxrwxrwx  1 noone  nogroup      0 Sep 01 06:22 new
drwxrwxrwx  1 noone  nogroup      0 Sep 02 06:33 new2
-rwxrwxrwx  1 noone  nogroup      283 Sep 02 06:35 pub
-rwxrwxrwx  1 noone  nogroup      283 Sep 02 06:36 puk
sftp-client>
```

Terminate the connection to the remote SFTP server.

```
sftp-client> quit
Bye
Connection closed.
<SwitchA>
```

RRPP CONFIGURATION

When configuring RRPP, go to these sections for information you are interested in:

- “RRPP Overview” on page 1139
- “RRPP Configuration Task List” on page 1146
- “Configuring Master Node” on page 1147
- “Configuring Transit Node” on page 1148
- “Configuring Edge Node” on page 1149
- “Configuring Assistant Edge Node” on page 1151
- “Displaying and Maintaining RRPP” on page 1152
- “RRPP Typical Configuration Examples” on page 1152

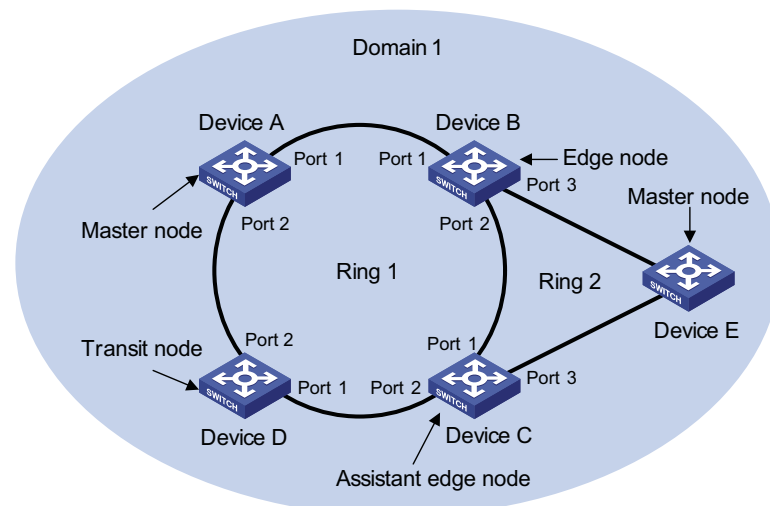
RRPP Overview

Rapid Ring Protection Protocol (RRPP) is an Ethernet ring-specific link layer protocol. It can not only prevent data loop from causing broadcast storm efficiently when the Ethernet ring is complete, but also restore communication channels among nodes on the Ethernet ring rapidly when a link is torn down.

Compared with Spanning Tree Protocol (STP), RRPP features:

- Expedited topology convergence
- Independent of the number of nodes on the Ethernet ring

Basic Concepts in RRPP **Figure 343** RRPP networking diagram



RRPP domain

The interconnected devices with the same domain ID and control VLANs constitute an RRPP domain. An RRPP domain contains multiple RRPP rings, in which one ring serves as the primary ring and other rings serve as sub rings. You can set a ring as either the primary ring or a sub ring.

As shown in Figure 343, Domain 1 is an RRPP domain, including two RRPP rings: Ring 1 and Ring 2. All the nodes on the two RRPP rings belong to the RRPP domain.

RRPP ring

A ring-shaped Ethernet topology is called an RRPP ring. An RRPP domain is built up on an RRPP ring. An RRPP ring falls into primary ring and sub ring. Both levels are set to 0 and 1 respectively when configuration.

As shown in Figure 343, Domain 1 contains two RRPP rings: Ring 1 and Ring 2. Ring 1 level is set to 0, meaning the primary ring; Ring 2 level is set to 1, meaning the sub ring.

For a ring, there are two cases:

- Health state: All the physical links on the Ethernet ring are connected.
- Disconnect state: Some physical link on the Ethernet ring fails.

Control VLAN and data VLAN

- Control VLAN is a VLAN specially designed to transfer RRPP packets. The ports accessing an RRPP ring on devices belong to the control VLAN of the ring and only these ports can join this VLAN. IP address configuration is prohibited on the ports of the control VLAN. You can configure a control VLAN for the primary ring (namely the primary control VLAN). However, the control VLAN of a sub ring (namely the secondary control VLAN) is assigned automatically by the system and its VLAN ID is the control VLAN ID of the primary ring plus 1.
- Data VLAN is a VLAN designed to transfer data packets, including the ports accessing the Ethernet ring and other ports on devices.

Node

Every device on an RRPP ring is referred to as a node. Node mode includes:

- Master node: Each ring has a master node primarily used to make loop detection and loop guard.
- Transit node: All the nodes excluding the master node on the primary ring; and all the nodes on a sub ring except for the master node and the nodes where the primary ring intersects with the sub ring.
- Edge node: A node residing on the primary ring and a sub ring at the same time. The node is a special transit node that serves as a transit node on the primary ring and an edge node on the sub ring.
- Assistant-edge node: A node residing on the primary ring and a sub ring at the same time. The node is a special transit node that serves as a transit node on the primary ring and an assistant-edge node on the sub ring. This node is used in conjunction with the edge node to detect the integrity of the primary ring and perform loop guard.

As shown in Figure 343, Ring 1 is the primary ring and Ring 2 is a sub ring. Device A is the master node of Ring 1, Device B, Device C and Device D are the transit nodes of Ring 1; Device E is the master node of Ring 2, Device B is the edge node of Ring 2, and Device C is the assistant edge node of Ring 2.

Primary port and secondary port

Each master node or transit node has two ports accessing an RRPP ring, in which one serves as the primary port and the other serves as the secondary port. You can determine the role of a port.

- 1 In terms of functionality, the difference between the primary port and the secondary port of a master node is:
 - The primary port and the secondary port are designed to play the role of sending and receiving loop-detect packets respectively.
 - When an RRPP ring is in health state, the secondary port of the master node will logically deny data VLANs and permit only the packets of the control VLANs.
 - When an RRPP ring is in disconnect state, the secondary port of the master node will permit data VLANs, that is, forward packets of data VLANs.
- 2 In terms of functionality, there is no difference between the primary port and the secondary port of the transit node. Both are designed for the transfer of protocol packets and data packets over an RRPP ring.

As shown in Figure 343, Device A is the master node of Ring 1. Port 1 and port 2 are the primary port and the secondary port of the master node on Ring 1 respectively. Device B, Device C and Device D are the transit nodes of Ring 1. Their port 1 and port 2 are the primary port and the secondary port on Ring 1 respectively.

Common port and edge port

Each edge node or assistant edge node have two ports accessing a sub ring, with one being a common port and the other being an edge port. Common port is a port accessing the primary ring and a sub ring simultaneously; and edge port is a port accessing only a sub ring.

As shown in Figure 343, Device B and Device C lie on Ring 1 and Ring 1. Device B's port 2 and Device C's port 1 access the primary ring and a sub ring at the same time, so they are common ports. Device B's port 3 and Device C's port 3 access only a sub ring, so they are edge ports.

Multi-domain intersection common port

Of the two ports on a node where rings of different domains intersect, the common port is the one on the primary ring that belongs to different domains at the same time. This port must not be on a sub ring. The role of the port is determined by user configuration.

Timers

The master node uses two timers to send and receive RRPP packets: the Hello timer and the Fail timer.

- The Hello timer is used for the primary port to send Health packets.

- The Fail timer is used for the secondary port to receive Health packets from the master node.

If the secondary port receives the Health packets before the Fail timer expires, the overall ring is in health state. Otherwise, the ring transits into disconnect state until the secondary port receives the Health packet again.



- *In an RRPP domain, a transit node learns the Hello timer value and the Fail timer value on the master node through the received Health packets, guaranteeing the consistency of two timer values across a ring.*
- *The Fail timer value must be greater than or equal to 3 times of the Hello timer value.*

RRPP Packets Table 89 shows the types of RRPP packets and their functions.

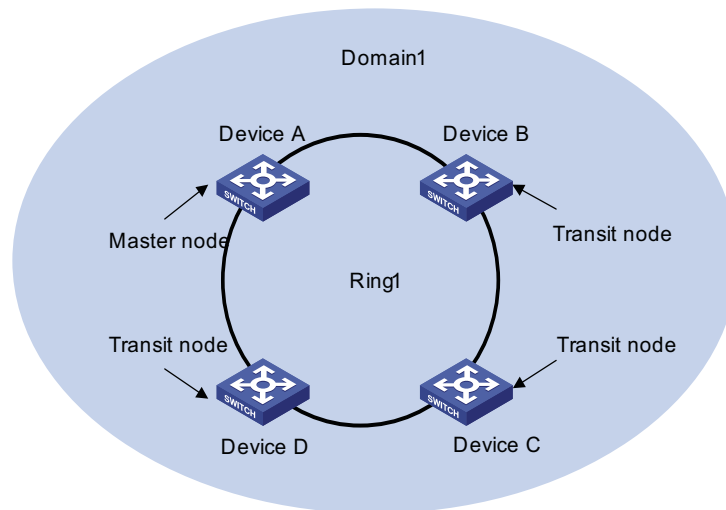
Table 89 RRPP packet types and their functions

Type	Description
Health	The master node initiates Health packets to detect the integrity of a ring in a network.
Link-Down	The transit node, the edge node or the assistant edge node initiates Link-Down packets to notify the master node the disappearance of a ring in case of a link failure.
Common-Flush-FDB	The master node initiates Common-Flush-FDB packets to notify the transit nodes to update their own MAC entries and ARP entries when an RRPP ring transits to disconnect state.
Complete-Flush-FDB	The master node initiates Complete-Flush-FDB packets to notify the transit nodes to update their own MAC entries and ARP entries, and release from blocking ports temporarily when an RRPP ring transits into health state.
Edge-Hello	The edge node initiates Edge-Hello packets to examine the links of the primary ring between the edge node and the assistant edge node.
Major-Fault	Assistant edge node initiates Major-Fault packets to notify the edge node of a failure when a link of primary ring between edge node and assistant edge node is torn down.

Typical RRPP Networking Here are several typical networking applications.

Single ring

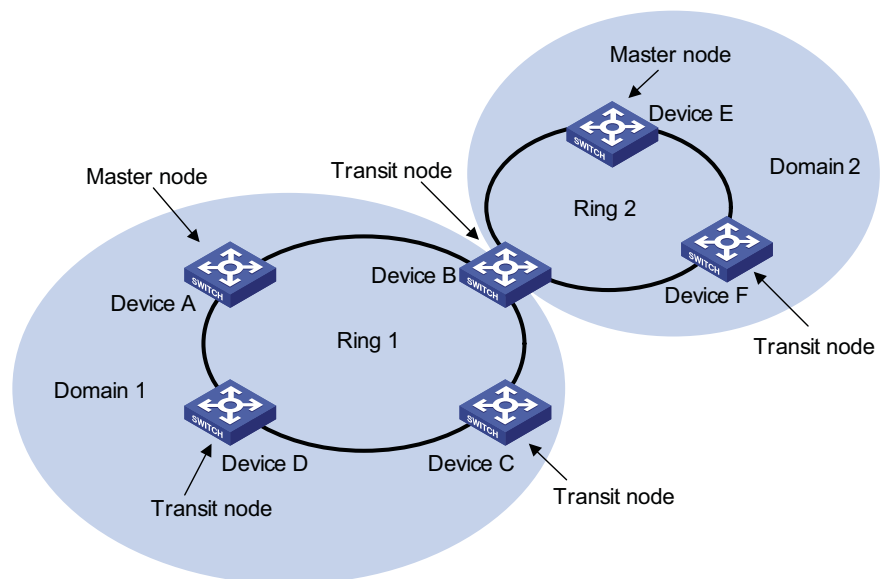
Figure 344 Single ring



There is only a single ring in the network topology. In this case, you only need to define an RRPP domain.

Multi-domain tangent rings

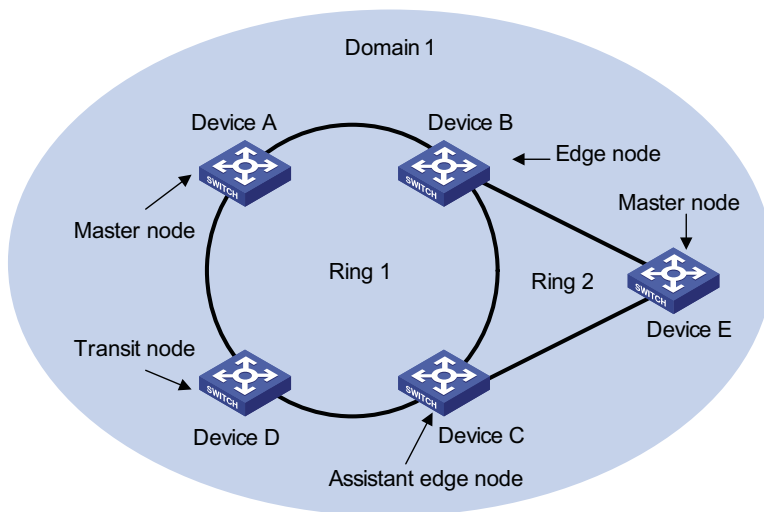
Figure 345 Multi-domain tangent rings



There are two or more rings in the network topology and only one common node between rings. In this case, you need define an RRPP domain for each ring.

Single-domain intersecting rings

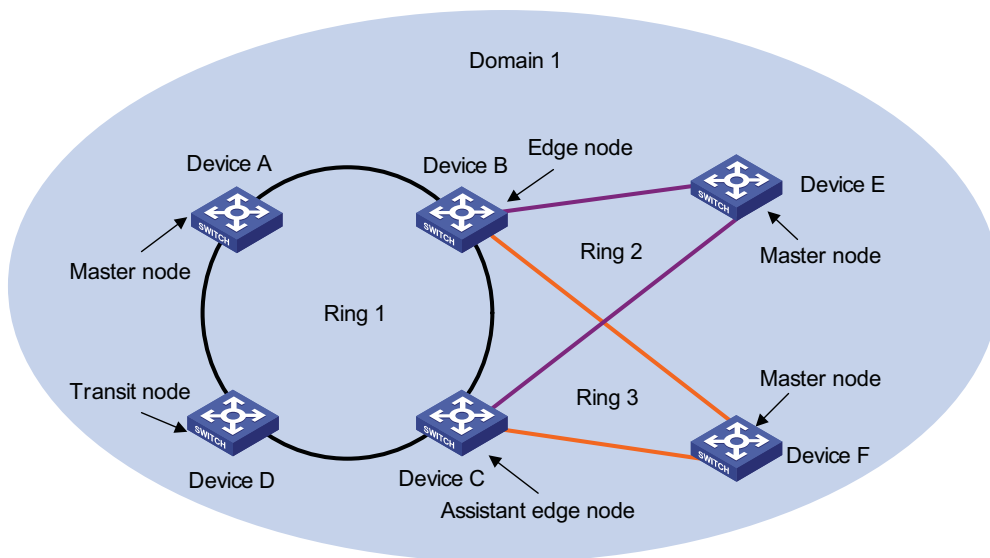
Figure 346 Single-domain intersecting rings



There are two or more rings in the network topology and two common nodes between rings. In this case, you only need to define an RRPP domain, and set one ring as the primary ring and other rings as sub rings.

Dual homed rings

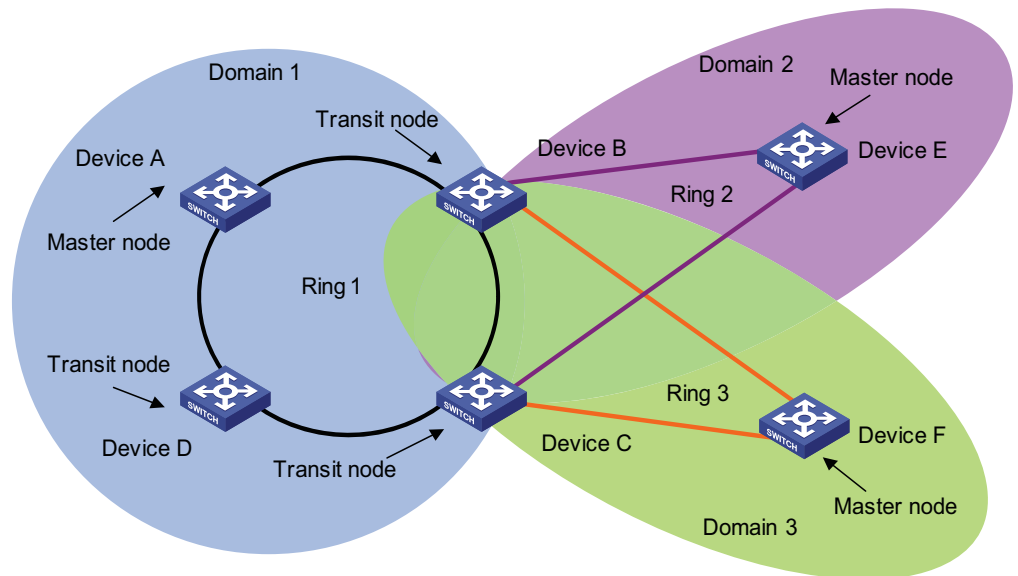
Figure 347 Dual homed rings



There are two or more rings in the network topology and two similar common nodes between rings. In this case, you only need to define an RRPP domain, and set one ring as the primary ring and other rings as sub rings.

Multi-domain intersecting rings

Figure 348 Multi-domain intersecting rings



There are two or more domains in a network, and there two different common nodes between any two domains. Figure 348 defines three RRPP domains, each containing one and only one RRPP primary ring. In the case of multi-domain intersection, the rings in different domains are independently configured. Each single domain can contain multiple rings, among which there must be one and only one primary ring. The data VLAN in one domain must be isolated from the data VLAN in another.

How RRPP Works **Polling mechanism**

The primary port of the master node sends Health packets across the control VLAN periodically.

- If the ring works properly, the secondary port of the master node will receive Health packets and the master node will maintain it in block state.
- If the ring is torn down, the secondary port of the master node will not receive Health packets after the timeout timer expires. The master node will release the secondary port from blocking data VLAN while sending Common-Flush-FDB packets to notify all transit nodes to update their own MAC entries and ARP entries.

Link down alarm mechanism

The transit node, the edge node or the assistant edge node sends Link-Down packets to the master node immediately when they find any port belonging to an RRPP domain is down. Upon the receipt of a Link-Down packet, the master node releases the secondary port from blocking data VLAN while sending Common-Flush-FDB packet to notify all the transit nodes, the edge nodes and the assistant nodes to update their own MAC entries and ARP entries.

Ring recovery

The master node may find the ring is restored after a period of time after the ports belonging to the RRPP domain on the transit node, the edge node or the assistant

edge node are up again. A temporary loop may arise in the data VLAN in this period. As a result, broadcast storm occurs.

To prevent temporary loops, non-master nodes block them immediately (and permits only the packets of the control VLAN) when they find their ports accessing the ring are up again. The blocked ports are activated only when the nodes ensure that no loop will be brought forth by these ports.

Broadcast storm suppression mechanism in a multi-homed sub ring in case of primary ring link failure

As shown in Figure 347, Ring 1 is the primary ring, and Ring 2 and Ring 3 are sub rings. When two links of the primary ring between the edge node and the assistant edge node are down, the master nodes of Ring 2 and Ring 3 will open their respective secondary ports, and thus a loop among B, C, E and F is generated. As a result, broadcast storm occurs.

In this case, to prevent from generating this loop, the edge node will block the edge port temporarily. The blocked edge port is activated only when the edge node ensures that no loop will be brought forth when the edge port is activated.

Protocols and Standards Related standard: RFC 3619.

RRPP Configuration Task List

Complete the following tasks to configure RRPP

Task	Description
"Configuring Master Node" on page 1147	Required
"Configuring Transit Node" on page 1148	Optional
"Configuring Edge Node" on page 1149	Optional
"Configuring Assistant Edge Node" on page 1151	Optional




CAUTION:

- *It is recommended to configure the primary ring first and then the sub ring when you configure an RRPP domain. Moreover, a Ring ID cannot be applied to more than one RRPP ring in one RRPP domain.*
- *If a device lies on multiple RRPP rings in an RRPP domain, only one primary ring exists. The device serves as either an edge node or an assistant edge node on the sub rings.*
- *The total number of rings configured on a device in all RRPP domains should not be greater than 16.*
- *Modification of node mode, port role and ring level of an RRPP ring is prohibited after configuration. If needed, you must first delete the existing configuration.*
- *The secondary port on the master node and a port on a sub ring node must not be configured as a multi-domain intersection common port, and the two ports that access the same node to the same RRPP ring must not be configured as multi-domain intersection common ports at the same time.*

- When configuring multi-domain intersecting rings, do not enable or disable the RRPP ring on which the multi-domain intersection common port resides with RRPP globally enabled.
- In the case of multi-domain intersection, the rings in different domains are independently configured. Each single domain can contain multiple rings, among which there must be one and only one primary ring. The data VLAN in one domain must be isolated from the data VLAN in another.

The ports accessing an RRPP ring must conform to the following conditions:

- Trunk port;
 - Layer 2 GE port;
 - Except for aggregation port and loopback port;
 - Port with STP, QinQ, 802.1x, MAC address authentication, voice VLAN disabled;
 - Do not enable OAM remote loopback function on an RRPP port. Otherwise, this may cause temporary broadcast storm;
 - Enable link status rapid report function on a port accessing an RRPP ring (the link-delay of the port is set to 0) to accelerate topology convergence.
-  ■ If you need to transparently transmit RRPP packets on a device without enabling RRPP, you should ensure only the two ports accessing an RRPP ring permits the packets of the control VLAN. Otherwise, the packets from other VLANs may go into the control VLAN in transparent transmission mode and strike the RRPP ring.
- Do not set the default VLAN ID of a port accessing an RRPP ring to primary control VLAN ID or secondary control VLAN ID (the latter is equal to the former plus 1), avoiding the influence on the proper receiving/sending of RRPP packets.

Configuring Master Node

Configuration Procedure Follow these steps to configure master node:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create an RRPP domain and enter its view	rrpp domain <i>domain-id</i>	Required
Specify control VLAN for the RRPP domain	control-vlan <i>vlan-id</i>	Required
Specify the current device as the master node of the ring, and specify the primary port and the secondary port	ring <i>ring-id</i> node-mode master [primary-port <i>interface-type interface-number</i>] [secondary-port <i>interface-type interface-number</i>] level <i>level-value</i>	Required

To do...	Use the command...	Remarks
Configure the timer for the RRPP domain	timer hello-timer <i>hello-value</i> fail-timer <i>fail-value</i>	Optional By default, the Hello timer value is 1 second and the Fail timer value is 3 seconds.
Enable the RRPP ring	ring <i>ring-id</i> enable	Required By default, the RRPP ring is disabled.
Return to system view	quit	-
Enable RRPP	rrpp enable	Required By default, RRPP is disabled.

**CAUTION:**

- The control VLAN configured for an RRPP domain must be a new one.
- Control VLAN configuration is required for configuring an RRPP ring.
- To use the **undo rrpp domain** command to remove an RRPP domain, you must ensure the RRPP domain has no RRPP ring.

Master Node Configuration Example

Network requirements

- Specify the device in RRPP domain 1;
- Set VLAN 4092 as the control VLAN;
- Specify the device as the master node of primary ring 1 in RRPP domain 1, GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port;
- Set the Hello timer value to 2 seconds and the Fail timer value to 7 seconds.

Configuration procedure

```
<Sysname> system-view
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] link-delay 0
[Sysname-GigabitEthernet1/0/1] quit
[Sysname] interface gigabitethernet 1/0/2
[Sysname-GigabitEthernet1/0/2] link-delay 0
[Sysname-GigabitEthernet1/0/2] quit
[Sysname] rrpp domain 1
[Sysname-rrpp-domain1] control-vlan 4092
[Sysname-rrpp-domain1] ring 1 node-mode master primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Sysname-rrpp-domain1] timer hello-timer 2 fail-timer 7
[Sysname-rrpp-domain1] ring 1 enable
[Sysname-rrpp-domain1] quit
[Sysname] rrpp enable
```

Configuring Transit Node

Configuration Procedure Follow these steps to configure transit node:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create an RRPP domain and enter its view	rrpp domain <i>domain-id</i>	Required

To do...	Use the command...	Remarks
Specify a control VLAN for the RRPP domain	control-vlan <i>vlan-id</i>	Required
Specify the current device as the transit node of the ring, and specify the primary port and the secondary port	ring <i>ring-id</i> node-mode transit [primary-port <i>interface-type interface-number</i>] [secondary-port <i>interface-type interface-number</i>] level <i>level-value</i>	Required
Enable the RRPP ring	ring <i>ring-id</i> enable	Required By default, the RRPP ring is disabled.
Return to system view	quit	-
Enable RRPP	rrpp enable	Required By default, RRPP is disabled.

**CAUTION:**

- The control VLAN configured for an RRPP domain must be a new one.
- Control VLAN configuration is required for configuring an RRPP ring.
- To use the **undo rrpp domain** command to remove an RRPP domain, you must ensure the RRPP domain has no RRPP ring.

Transit Node Configuration Example

Network requirements

- Specify the device in RRPP domain 1;
- Set VLAN 4092 as the control VLAN;
- Specify the device as the transit node of primary ring 1 in RRPP domain 1, GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port.

Configuration procedure

```
<Sysname> system-view
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] link-delay 0
[Sysname-GigabitEthernet1/0/1] quit
[Sysname] interface gigabitethernet 1/0/2
[Sysname-GigabitEthernet1/0/2] link-delay 0
[Sysname-GigabitEthernet1/0/2] quit
[Sysname] rrpp domain 1
[Sysname-rrpp-domain1] control-vlan 4092
[Sysname-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Sysname-rrpp-domain1] ring 1 enable
[Sysname-rrpp-domain1] quit
[Sysname] rrpp enable
```

Configuring Edge Node

Configuration Procedure

Follow these steps to configure edge node:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Create an RRPP domain and enter its view	rrpp domain <i>domain-id</i>	Required
Specify a control VLAN for the RRPP domain	control-vlan <i>vlan-id</i>	Required
Specify the current device as the transit node of the primary ring, and specify the primary port and the secondary port	ring <i>ring-id</i> node-mode transit [primary-port <i>interface-type interface-number</i>] [secondary-port <i>interface-type interface-number</i>] level <i>level-value</i>	Required
Specify the current device as the edge node of a sub ring, and specify the common port and the edge port	ring <i>ring-id</i> node-mode edge [common-port <i>interface-type interface-number</i>] [edge-port <i>interface-type interface-number</i>]	Required
Enable the primary ring	ring <i>ring-id</i> enable	Required By default, the RRPP ring is disabled.
Enable the sub ring	ring <i>ring-id</i> enable	Required By default, the RRPP ring is disabled.
Return to system view	quit	-
Enable RRPP	rrpp enable	Required By default, RRPP is disabled.

**CAUTION:**

- *The control VLAN configured for an RRPP domain must be a new one.*
- *Control VLAN configuration is required for configuring an RRPP ring.*
- *A Ring ID cannot be applied to more than one RRPP ring in an RRPP domain.*
- *You must first configure the primary ring and then the sub ring when configuring an edge node. Moreover, you must remove all sub ring configurations before deleting the primary ring configuration of an edge node. However, the RRPP ring enabled cannot be deleted.*
- *To use the **undo rrpp domain** command to remove an RRPP domain, you must ensure the RRPP domain has no RRPP ring.*

Edge Node Configuration Example

Networking requirements

- Specify the device in RRPP domain 1;
- Set VLAN 4092 as the control VLAN;
- Specify the device as the transit node of primary ring 1 in RRPP domain 1, GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port;
- Specify the device as the edge node of sub ring 2 in RRPP domain 1, GigabitEthernet 1/0/2 as a common port and GigabitEthernet 1/0/4 as an edge port.

Configuration procedure

```
<Sysname> system-view
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] link-delay 0
[Sysname-GigabitEthernet1/0/1] quit
[Sysname] interface gigabitethernet 1/0/2
[Sysname-GigabitEthernet1/0/2] link-delay 0
[Sysname-GigabitEthernet1/0/2] quit
[Sysname] interface gigabitethernet 1/0/4
[Sysname-GigabitEthernet1/0/4] link-delay 0
[Sysname-GigabitEthernet1/0/4] quit
[Sysname] rrpp domain 1
[Sysname-rrpp-domain1] control-vlan 4092
[Sysname-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Sysname-rrpp-domain1] ring 2 node-mode edge common-port gigabitethernet 1/0/2 edge-port gigabitethernet 1/0/4
[Sysname-rrpp-domain1] ring 1 enable
[Sysname-rrpp-domain1] ring 2 enable
[Sysname-rrpp-domain1] quit
[Sysname] rrpp enable
```

Configuring Assistant Edge Node

Configuration Procedure Follow these steps to configure assistant edge node:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create an RRPP domain and enter its view	rrpp domain <i>domain-id</i>	Required
Specify a control VLAN for the RRPP domain	control-vlan <i>vlan-id</i>	Required
Specify the current device as the transit node of the primary ring, and specify the primary port and the secondary port	ring <i>ring-id</i> node-mode transit [primary-port <i>interface-type interface-number</i>] [secondary-port <i>interface-type interface-number</i>] level <i>level-value</i>	Required
Specify the current device as the assistant edge node of the sub ring, and specify a common port and an edge port	ring <i>ring-id</i> node-mode assistant-edge [common-port <i>interface-type interface-number</i>] [edge-port <i>interface-type interface-number</i>]	Required
Enable the primary ring	ring <i>ring-id</i> enable	Required By default, the RRPP ring is disabled.
Enable the sub ring	ring <i>ring-id</i> enable	Required By default, the RRPP ring is disabled.
Return to system view	quit	-
Enable RRPP	rrpp enable	Required By default, RRPP is disabled.



CAUTION:

- The control VLAN configured for an RRPP domain must be a new one.

- Control VLAN configuration is required for configuring an RRPP ring.
- A Ring ID cannot be applied to more than on RRPP ring in an RRPP domain.
- You must first configure the primary ring and then the sub ring when configuring an edge node. Moreover, you must remove all sub ring configurations before deleting the primary ring configuration of an edge node. However, the RRPP ring enabled cannot be deleted.
- To use the **undo rrpp domain** command to remove an RRPP domain, you must ensure the RRPP domain has no RRPP ring.

Assistant Edge Node Configuration Example

Networking requirements

- Specify the device in RRPP domain 1;
- Set VLAN 4092 as the control VLAN;
- Specify the device as the transit node of primary ring 1 in RRPP domain 1, GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port;
- Specify the device as the assistant edge node of sub ring 2 in RRPP domain 1, GigabitEthernet 1/0/2 as the common port and GigabitEthernet 1/0/4 as the edge port.

Configuration procedure

```
<Sysname> system-view
[Sysname] interface gigabitethernet 1/0/1
[Sysname-GigabitEthernet1/0/1] link-delay 0
[Sysname-GigabitEthernet1/0/1] quit
[Sysname] interface gigabitethernet 1/0/2
[Sysname-GigabitEthernet1/0/2] link-delay 0
[Sysname-GigabitEthernet1/0/2] quit
[Sysname] interface gigabitethernet 1/0/4
[Sysname-GigabitEthernet1/0/4] link-delay 0
[Sysname-GigabitEthernet1/0/4] quit
[Sysname] rrpp domain 1
[Sysname-rrpp-domain1] control-vlan 4092
[Sysname-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Sysname-rrpp-domain1] ring 2 node-mode assistant-edge common-port gigabitethernet 1/0/2 edge-port gigabitethernet 1/0/4
[Sysname-rrpp-domain1] ring 1 enable
[Sysname-rrpp-domain1] ring 2 enable
[Sysname-rrpp-domain1] quit
[Sysname] rrpp enable
```

Displaying and Maintaining RRPP

To do...	Use the command...	Remarks
Display brief information about RRPP configuration	display rrpp brief	Available in any view
Display detailed information about RRPP configuration	display rrpp verbose domain <i>domain-id</i> [ring <i>ring-id</i>]	
Display RRPP statistics	display rrpp statistics domain <i>domain-id</i> [ring <i>ring-id</i>]	
Clear RRPP statistics	reset rrpp statistics domain <i>domain-id</i> [ring <i>ring-id</i>]	Available in user view

RRPP Typical Configuration Examples

This section covers these topics:

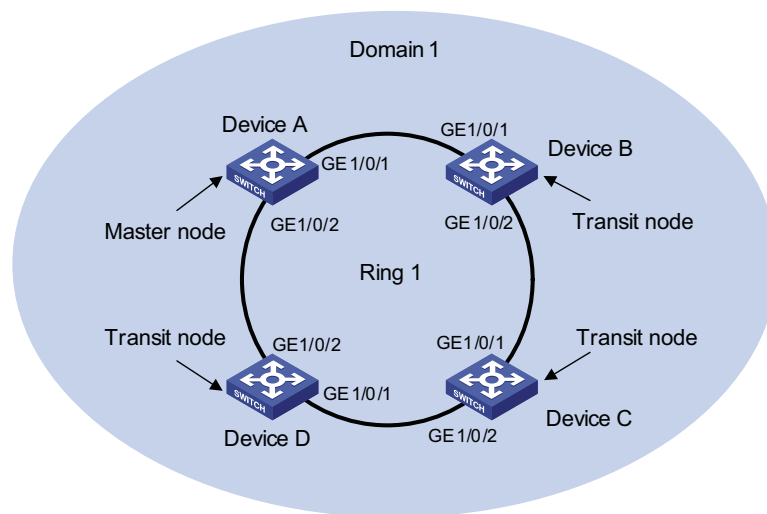
- Configuring Single Ring Topology
- Configuring Intersecting Ring Topology

Configuring Single Ring Topology

Networking requirements

- Device A, Device B, Device C and Device D constitute RRPP domain 1;
- Specify the control VLAN of RRPP domain 1 as VLAN 4092;
- Device A, Device B, Device C and Device D constitute primary ring 1;
- Specify Device A as the master node of primary ring 1, GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port;
- Specify Device B, Device C and Device D as the transit nodes of primary ring 1, their GigabitEthernet 1/0/1 as the primary port and GigabitEthernet 1/0/2 as the secondary port;
- The timers of the primary ring adopt the default value.

Figure 349 Single ring networking diagram



Configuration considerations

First, determine the node mode of a device in an RRPP ring, and then perform the following configurations on a per-device basis:

- Create an RRPP domain.
- Specify the control VLAN for the RRPP domain.
- Specify the node mode of a device on the primary ring and the ports accessing the RRPP ring on the device.
- Enable the RRPP ring.
- Enable RRPP

Configuration procedure

- 1 Perform the following configuration on Device A:

```
<Device A> system-view
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] link-delay 0
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] link-delay 0
[DeviceA-GigabitEthernet1/0/2] quit
[Device A] rrpp domain 1
[Device A-rrpp-domain1] control-vlan 4092
[Device A-rrpp-domain1] ring 1 node-mode master primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Device A-rrpp-domain1] ring 1 enable
```

```
[Device A-rrpp-domain1] quit
[Device A] rrpp enable
```

2 Perform the following configuration on Device B:

```
<Device B> system-view
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] link-delay 0
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] link-delay 0
[DeviceB-GigabitEthernet1/0/2] quit
[Device B] rrpp domain 1
[Device B-rrpp-domain1] control-vlan 4092
[Device B-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Device B-rrpp-domain1] ring 1 enable
[Device B-rrpp-domain1] quit
[Device B] rrpp enable
```

3 Perform the following configuration on Device C:

```
<Device C> system-view
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] link-delay 0
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] link-delay 0
[DeviceC-GigabitEthernet1/0/2] quit
[Device C] rrpp domain 1
[Device C-rrpp-domain1] control-vlan 4092
[Device C-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Device C-rrpp-domain1] ring 1 enable
[Device C-rrpp-domain1] quit
[Device C] rrpp enable
```

4 Perform the following configuration on Device D:

```
<Device D> system-view
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] link-delay 0
[DeviceD-GigabitEthernet1/0/1] quit
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] link-delay 0
[DeviceD-GigabitEthernet1/0/2] quit
[Device D] rrpp domain 1
[Device D-rrpp-domain1] control-vlan 4092
[Device D-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Device D-rrpp-domain1] ring 1 enable
[Device D-rrpp-domain1] quit
[Device D] rrpp enable
```

After the above configuration, you can use the **display** command to view RRPP configuration.

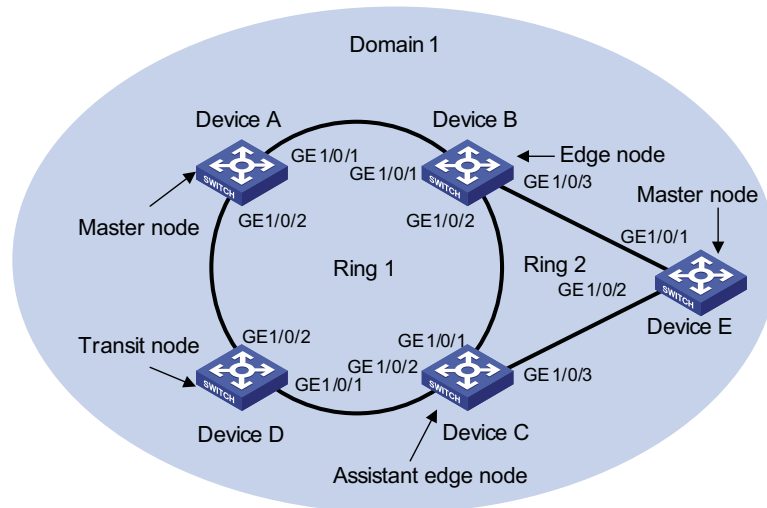
Configuring Single-Domain Intersecting Ring Topology

Networking requirements

- Device A, Device B, Device C and Device D constitute RRPP domain 1.
- VLAN 4092 is the control VLAN of RRPP domain 1;
- Device A, Device B, Device C and Device D constitute primary ring 1;
- Device B, Device C and Device E constitute sub ring 2;
- Device A is the master node of primary ring 1, GigabitEthernet 1/0/1 is the primary port and GigabitEthernet 1/0/2 is the secondary port;
- Device E is the master node of sub ring 2, GigabitEthernet 1/0/1 is the primary port and GigabitEthernet 1/0/2 is the secondary port;
- Device B is the transit node of primary ring 1 and the edge node of sub ring 2, GigabitEthernet 1/0/2 is the common port and GigabitEthernet 1/0/3 is the edge port;
- Device C is the transit node of primary ring 1 and the assistant edge node of sub ring 1, GigabitEthernet 1/0/1 is the common port and GigabitEthernet 1/0/3 is the edge port;

- Device D is the transit node of primary ring 1, GigabitEthernet 1/0/1 is the primary port and GigabitEthernet 1/0/2 is the secondary port;
- The timers of both the primary ring and the sub ring adopt the default value.

Figure 350 Networking diagram for single-domain intersecting rings configuration



Configuration considerations

First, determine the primary ring and sub ring in an RRPP domain, node mode of a device on each RRPP ring, and then perform the following configuration on a per-device basis:

- Create an RRPP domain.
- Specify the control VLAN for the RRPP domain.
- Specify the node mode of a device on an RRPP ring and the ports accessing the RRPP ring on the device.
- Enable these two RRPP rings.
- Enable RRPP

Configuration procedure

- 1 Perform the following configuration on Device A:

```
<Device A> system-view
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] link-delay 0
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] link-delay 0
[DeviceA-GigabitEthernet1/0/2] quit
[Device A] rrpp domain 1
[Device A-rrpp-domain1] control-vlan 4092
[Device A-rrpp-domain1] ring 1 node-mode master primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Device A-rrpp-domain1] ring 1 enable
[Device A-rrpp-domain1] quit
[Device A] rrpp enable
```

- 2 Perform the following configuration on Device B:

```
<Device B> system-view
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] link-delay 0
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] link-delay 0
[DeviceB-GigabitEthernet1/0/2] quit
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] link-delay 0
```

```
[DeviceB-GigabitEthernet1/0/3] quit
[Device B] rrpp domain 1
[Device B-rrpp-domain1] control-vlan 4092
[Device B-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Device B-rrpp-domain1] ring 2 node-mode edge common-port gigabitethernet 1/0/2 edge-port gigabitethernet 1/0/3
[Device B-rrpp-domain1] ring 1 enable
[Device B-rrpp-domain1] ring 2 enable
[Device B-rrpp-domain1] quit
[Device B] rrpp enable
```

3 Perform the following configuration on Device C:

```
<Device C> system-view
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] link-delay 0
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] link-delay 0
[DeviceC-GigabitEthernet1/0/2] quit
[DeviceC] interface gigabitethernet 1/0/3
[DeviceC-GigabitEthernet1/0/3] link-delay 0
[DeviceC-GigabitEthernet1/0/3] quit
[Device C] rrpp domain 1
[Device C-rrpp-domain1] control-vlan 4092
[Device C-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Device C-rrpp-domain1] ring 2 node-mode assistant-edge common-port gigabitethernet 1/0/1 edge-port gigabitethernet 1/0/3
[Device C-rrpp-domain1] ring 1 enable
[Device C-rrpp-domain1] ring 2 enable
[Device C-rrpp-domain1] quit
[Device C] rrpp enable
```

4 Perform the following configuration on Device D:

```
<Device D> system-view
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] link-delay 0
[DeviceD-GigabitEthernet1/0/1] quit
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] link-delay 0
[DeviceD-GigabitEthernet1/0/2] quit
[Device D] rrpp domain 1
[Device D-rrpp-domain1] control-vlan 4092
[Device D-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Device D-rrpp-domain1] ring 1 enable
[Device D-rrpp-domain1] quit
[Device D] rrpp enable
```

5 Perform the following configuration on Device E:

```
<Device E> system-view
[DeviceE] interface gigabitethernet 1/0/1
[DeviceE-GigabitEthernet1/0/1] link-delay 0
[DeviceE-GigabitEthernet1/0/1] quit
[DeviceE] interface gigabitethernet 1/0/2
[DeviceE-GigabitEthernet1/0/2] link-delay 0
[DeviceE-GigabitEthernet1/0/2] quit
[Device E] rrpp domain 1
[Device E-rrpp-domain1] control-vlan 4092
[Device E-rrpp-domain1] ring 2 node-mode master primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 1
[Device E-rrpp-domain1] ring 2 enable
[Device E-rrpp-domain1] quit
[Device E] rrpp enable
```

After the configuration, you can use the **display** command to view RRPP configuration result.

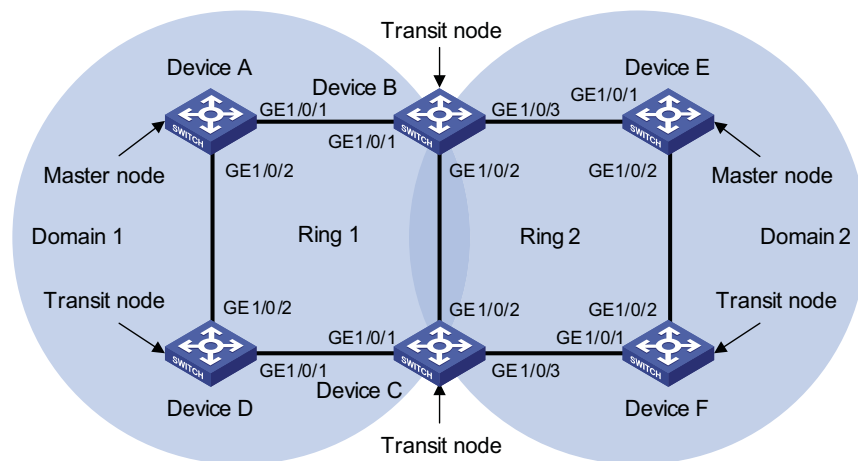
Configuring Multi-Domain Intersecting Ring Topology

Networking requirements

- Device A, Device B, Device C and Device D constitute RRPP domain 1, and Device E, Device F, Device C and Device B constitute RRPP domain 2;
- VLAN 4090 is the control VLAN of RRPP domain 1, and VLAN 4092 is the control VLAN of RRPP domain 2;
- Device A, Device B, Device C and Device D constitute primary ring 1;
- Device E, Device F, Device C and Device B constitute primary ring 2;

- On primary ring 1 in RRPP domain 1, Device A is the master node, GigabitEthernet 1/0/1 is the primary port and GigabitEthernet 1/0/2 is the secondary port;
- On primary ring 2 in RRPP domain 2, Device E is the master node, GigabitEthernet 1/0/1 is the primary port and GigabitEthernet 1/0/2 is the secondary port;
- Device B is a transit node on primary ring 1 in RRPP domain 1 and a transit node on primary ring 2 in RRPP domain 2, GigabitEthernet 1/0/2 is a multi-domain intersection common port;
- Device C is a transit node on primary ring 1 in RRPP domain 1 and a transit node on primary ring 2 in RRPP domain 2, and GigabitEthernet 1/0/2 is a multi-domain intersection common port;
- Device D is a transit node on primary ring 1 in RRPP domain 1, GigabitEthernet 1/0/1 is the primary port and GigabitEthernet 1/0/2 is the secondary port;
- Device F is a transit node on primary ring 2 in RRPP domain 2, GigabitEthernet 1/0/1 is the primary port and GigabitEthernet 1/0/2 is the secondary port;
- Use default values for timers on the primary ring in each domain.

Figure 351 Networking diagram for multi-domain intersecting ring configuration



Configuration considerations

First, determine the node roles on the primary ring and sub ring in each RRPP domain, and then perform the following configuration domain by domain on each device:

- Create RRPP domains;
- Specify the control VLAN for each domain;
- Specify the node role of each device on each RRPP ring and the ports that access the device to the RRPP rings;
- Enable the RRPP rings;
- Enable RRPP after completing the configuration of both domains.

Configuration procedure

1 Perform the following configuration on Device A:

```
<Device A> system-view
[DeviceA] interface gigabitethernet 1/0/1
[DeviceA-GigabitEthernet1/0/1] link-delay 0
[DeviceA-GigabitEthernet1/0/1] quit
[DeviceA] interface gigabitethernet 1/0/2
[DeviceA-GigabitEthernet1/0/2] link-delay 0
[DeviceA-GigabitEthernet1/0/2] quit
[Device A] rrpp domain 1
[Device A-rrpp-domain1] control-vlan 4090
[Device A-rrpp-domain1] ring 1 node-mode master primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Device A-rrpp-domain1] ring 1 enable
[Device A-rrpp-domain1] quit
[Device A] rrpp enable
```

2 Perform the following configuration on Device B:

```
<Device B> system-view
[DeviceB] interface gigabitethernet 1/0/1
[DeviceB-GigabitEthernet1/0/1] link-delay 0
[DeviceB-GigabitEthernet1/0/1] quit
[DeviceB] interface gigabitethernet 1/0/2
[DeviceB-GigabitEthernet1/0/2] link-delay 0
[DeviceB-GigabitEthernet1/0/2] quit
[DeviceB] interface gigabitethernet 1/0/3
[DeviceB-GigabitEthernet1/0/3] link-delay 0
[DeviceB-GigabitEthernet1/0/3] quit
[Device B] rrpp domain 1
[Device B-rrpp-domain1] control-vlan 4090
[Device B-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Device B-rrpp-domain1] ring 1 enable
[Device B-rrpp-domain1] quit
[Device B] rrpp domain 2
[Device B-rrpp-domain2] control-vlan 4092
[Device B-rrpp-domain2] ring 2 node-mode transit primary-port gigabitethernet 1/0/2 secondary-port gigabitethernet 1/0/3 level 0
[Device B-rrpp-domain2] ring 2 enable
[Device B-rrpp-domain2] quit
[Device B] rrpp enable
```

3 Perform the following configuration on Device C:

```
<Device C> system-view
[DeviceC] interface gigabitethernet 1/0/1
[DeviceC-GigabitEthernet1/0/1] link-delay 0
[DeviceC-GigabitEthernet1/0/1] quit
[DeviceC] interface gigabitethernet 1/0/2
[DeviceC-GigabitEthernet1/0/2] link-delay 0
[DeviceC-GigabitEthernet1/0/2] quit
[DeviceC] interface gigabitethernet 1/0/3
[DeviceC-GigabitEthernet1/0/3] link-delay 0
[DeviceC-GigabitEthernet1/0/3] quit
[Device C] rrpp domain 1
[Device C-rrpp-domain1] control-vlan 4090
[Device C-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Device C-rrpp-domain1] ring 1 enable
[Device C-rrpp-domain1] quit
[Device C] rrpp domain 2
[Device C-rrpp-domain2] control-vlan 4092
[Device C-rrpp-domain2] ring 2 node-mode transit primary-port gigabitethernet 1/0/3 secondary-port gigabitethernet 1/0/2 level 0
[Device C-rrpp-domain2] ring 2 enable
[Device C-rrpp-domain2] quit
[Device C] rrpp enable
```

4 Perform the following configuration on Device D:

```
<Device D> system-view
[DeviceD] interface gigabitethernet 1/0/1
[DeviceD-GigabitEthernet1/0/1] link-delay 0
[DeviceD-GigabitEthernet1/0/1] quit
[DeviceD] interface gigabitethernet 1/0/2
[DeviceD-GigabitEthernet1/0/2] link-delay 0
[DeviceD-GigabitEthernet1/0/2] quit
[Device D] rrpp domain 1
[Device D-rrpp-domain1] control-vlan 4090
[Device D-rrpp-domain1] ring 1 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Device D-rrpp-domain1] ring 1 enable
[Device D-rrpp-domain1] quit
[Device D] rrpp enable
```

5 Perform the following configuration on Device E:

```
<Device E> system-view
[DeviceE] interface gigabitethernet 1/0/1
[DeviceE-GigabitEthernet1/0/1] link-delay 0
[DeviceE-GigabitEthernet1/0/1] quit
```

```
[DeviceE] interface gigabitethernet 1/0/2
[DeviceE-GigabitEthernet1/0/2] link-delay 0
[DeviceE-GigabitEthernet1/0/2] quit
[Device E] rrpp domain 2
[Device E-rrpp-domain2] control-vlan 4092
[Device E-rrpp-domain2] ring 2 node-mode master primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Device E-rrpp-domain2] ring 2 enable
[Device E-rrpp-domain2] quit
[Device E] rrpp enable
```

6 Perform the following configuration on Device F:

```
<Device F> system-view
[DeviceF] interface gigabitethernet 1/0/1
[DeviceF-GigabitEthernet1/0/1] link-delay 0
[DeviceF-GigabitEthernet1/0/1] quit
[DeviceF] interface gigabitethernet 1/0/2
[DeviceF-GigabitEthernet1/0/2] link-delay 0
[DeviceF-GigabitEthernet1/0/2] quit
[Device F] rrpp domain 2
[Device F-rrpp-domain2] control-vlan 4092
[Device F-rrpp-domain2] ring 2 node-mode transit primary-port gigabitethernet 1/0/1 secondary-port gigabitethernet 1/0/2 level 0
[Device F-rrpp-domain2] ring 2 enable
[Device F-rrpp-domain2] quit
[Device F] rrpp enable
```

After the configuration, you can use the **display** command to view RRPP configuration result.

91

PORT SECURITY CONFIGURATION

When configuring port security, go to these sections for information you are interested in:

- "Introduction to Port Security" on page 1161
- "Port Security Configuration Task List" on page 1164
- "Displaying and Maintaining Port Security" on page 1169
- "Port Security Configuration Examples" on page 1169
- "Troubleshooting Port Security" on page 1178

Introduction to Port Security

Port Security Overview

Port security is a MAC address-based security mechanism for network access controlling. It is an extension to the existing 802.1x authentication and MAC authentication. It controls the access of unauthorized devices to the network by checking the source MAC address of an inbound frame and the access to unauthorized devices by checking the destination MAC address of an outbound frame.

With port security, you can define various port security modes to make a device learn only legal source MAC addresses, so that you can implement different network security management as needed. When a port security-enabled device detects an illegal frame, it triggers the corresponding port security feature and takes a pre-defined action automatically. This reduces your maintenance workload and greatly enhances system security.

The following types of frames are classified as illegal:

- Received frames with unknown source MAC addresses when MAC address learning is disabled.
- Received frames with unknown source MAC addresses when the number of MAC addresses learned by the port has already reached the upper limit.
- Frames from unauthenticated users.

Port Security Features NTK

The need to know (NTK) feature checks the destination MAC addresses in outbound frames and allows frames to be sent to only devices passing authentication, thus preventing illegal devices from intercepting network traffic.

Intrusion protection

The intrusion protection feature checks the source MAC addresses in inbound frames and takes a pre-defined action accordingly upon detecting illegal frames. The action may be disabling the port temporarily, disabling the port permanently, or blocking frames with the MAC address for three minutes (unmodifiable).

Trap

The trap feature enables the device to send trap messages upon detecting specified frames that result from, for example, intrusion or user login/logout operations, helping you monitor special activities.

Port Security Modes Table 90 details the port security modes.

Table 90 Port security modes

Security mode	Description	Features
noRestrictions	Port security is disabled on the port and access to the port is not restricted.	In this mode, neither the NTK nor the intrusion protection feature is triggered.
autoLearn	In this mode, a port can learn a specified number of MAC addresses and save those addresses as secure MAC addresses. It permits only frames whose source MAC addresses are secure MAC addresses or static MAC addresses configured by using the mac-address static command. When the number of secure MAC addresses reaches the upper limit, the port changes to work in secure mode.	In either mode, the device will trigger NTK and intrusion protection upon detecting an illegal frame.
secure	In this mode, a port is disabled from learning MAC addresses and permits only frames whose source MAC addresses are secure MAC addresses or static MAC addresses configured by using the mac-address static command.	
userLogin	In this mode, a port performs 802.1x authentication of users in portbased mode.	In this mode, neither NTK nor intrusion protection will be triggered.

Table 90 Port security modes

Security mode	Description	Features
userLoginSecure	In this mode, a port performs 802.1x authentication of users in portbased mode and services only one user passing 802.1x authentication.	In any of these modes, the device will trigger NTK and intrusion protection upon detecting an illegal frame.
userLoginWithOUI	Similar to the userLoginSecure mode, a port in this mode performs 802.1x authentication of users and services only one user passing 802.1x authentication. A MAC address being a specified OUI (organizationally unique identifier) are also allowed on the port.	
macAddressWithRadius	In this mode, a port performs MAC authentication of users.	
macAddressOrUserLoginSecure	This mode is the combination of the userLoginSecure and macAddressWithRadius modes, with 802.1x authentication having a higher priority. the port performs MAC authentication upon receiving non-802.1x frames and performs 802.1x authentication first upon receiving 802.1x frames. If 802.1x authentication fails, the port performs MAC authentication.	
macAddressElseUserLoginSecure	This mode is the combination of the macAddressWithRadius and userLoginSecure modes, with MAC authentication having a higher priority. <ul style="list-style-type: none"> ■ Upon receiving a non-802.1x frame, a port in this mode performs only MAC authentication. ■ Upon receiving an 802.1x frame, the port performs MAC authentication and then, if MAC authentication fails, 802.1x authentication. 	
userLoginSecureExt	In this mode, a port performs 802.1x authentication of users in macbased mode and supports multiple concurrent users.	
macAddressOrUserLoginSecureExt	This mode is similar to macAddressOrUserLoginSecure mode. The difference is that this mode allows a port to support multiple 802.1x and MAC authentication users.	
macAddressElseUserLoginSecureExt	This mode is similar to macAddressElseUserLoginSecure mode. The difference is that this mode allows a port to support multiple 802.1x and MAC authentication users.	



- *Currently, port security supports two authentication methods: 802.1x and MAC authentication. Different port security modes employ different authentication method or different combinations of authentication methods.*
- *The maximum number of authenticated users that a port can support is the smaller one between the maximum number of secure MAC addresses and the maximum number of concurrent users that the mode of the port supports.*

Port Security Configuration Task List

Complete the following tasks to configure port security:

Task	Remarks
"Enabling Port Security" on page 1164	Required
"Setting the Maximum Number of Secure MAC Addresses" on page 1165	Optional
"Setting the Port Security Mode" on page 1165	Required
"Configuring Port Security Features" on page 1167	"Configuring NTK" on page 1167 "Configuring Intrusion Protection" on page 1167 "Configuring Trapping" on page 1167
"Configuring Secure MAC Addresses" on page 1168	Optional
"Ignoring the Authorization Information from the Server" on page 1168	Optional

Enabling Port Security

Configuration Prerequisites

Before enabling port security, you need to disable 802.1x and MAC authentication globally.

Configuration Procedure

Follow these steps to enable port security:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable port security	port-security enable	Required Disabled by default



Enabling port security resets the following configurations on a port to the defaults bracketed, making them dependent completely on the port security mode:

- 802.1x (disabled), port access control method (macbased), and port access control mode (auto)
- MAC authentication (disabled)

Disabling port security resets the following configurations on a port to the defaults bracketed:

- Port security mode (noRestrictions)
- 802.1x (disabled), port access control method (macbased), and port access control mode (auto)
- MAC authentication (disabled)

Port security cannot be disabled if there is any user present on a port.



For configuration information about 802.1x authentication and MAC authentication, refer to "802.1x Configuration" on page 715 and "MAC Authentication Configuration" on page 739.

Setting the Maximum Number of Secure MAC Addresses

With port security enabled, more than one authenticated user is allowed on a port. The number of authenticated users allowed, however, cannot exceed the specified upper limit.

By setting the maximum number of secure MAC addresses allowed on a port, you can

- Control the maximum number of users who are allowed access the network through the port
- Control the number of secure MAC addresses that can be added with port security

This configuration is different from that of the maximum number of MAC addresses that can be learned by the port in MAC address management.

Follow these steps to set the maximum number of secure MAC addresses allowed on a port:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	-
Set the maximum number of secure MAC addresses allowed on a port	port-security max-mac-count <i>count-value</i>	Required Not limited by default

Setting the Port Security Mode

Before setting the port security mode, ensure that:

- 802.1x is disabled, the port access control method is macbased, and the port access control mode is auto.
- MAC authentication is disabled.

Otherwise, you will see an error message and your configuration will fail.

On the other hand, after setting the port security mode on a port, you cannot change any of the above configurations.



- *With port security disabled, you can configure the port security mode but your configuration does not take effect.*
- *With port security enabled, you can change the port security mode of a port only when the port is operating in noRestrictions mode, the default mode. You can use the **undo port-security port-mode** command to restore the default port security mode.*
- *You cannot change the port security mode of a port when any user is present on the port.*
- *Configuration of port security mode and aggregation are mutually exclusive. You cannot configure both of them on a port.*

Enabling the autoLearn Mode

Configuration prerequisites

Before enabling the autoLearn mode, you need to set the maximum number of secure MAC addresses allowed on the port.

Configuration procedure

Follow these steps to enable the autoLearn mode:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable the autoLearn mode	port-security port-mode autolearn	Required By default, a port operates in noRestrictions mode.



When a port operates in autoLearn mode, you cannot change the maximum number of secure MAC addresses allowed on the port.

Enabling the userLoginWithOUI Mode

In userLoginWithOUI mode, a port supports one 802.1x user as well as users whose MAC addresses have an OUI value among the specified ones.

Follow these steps to enable the userLoginWithOUI mode:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Set an OUI value for user authentication	port-security oui <i>oui-value</i> index <i>index-value</i>	Optional Not configured by default
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable the userLoginWithOUI mode	port-security port-mode userlogin-withoui	Required By default, a port operates in noRestrictions mode.



- *An organizationally unique identifier (OUI), the left-most 24 bits of a MAC address, is a globally unique identifier assigned by IEEE to a certain manufacturer.*
- *You can configure multiple OUI values.*

Enabling any other Port Security Mode

Follow these steps to enable any other port security mode:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i> view	-

To do...	Use the command...	Remarks
Set the port security mode	port-security port-mode { mac-authentication mac-else-userlogin-secure mac-else-userlogin-secure-ext secure userlogin userlogin-secure userlogin-secure-ext userlogin-secure-or-mac userlogin-secure-or-mac-ext }	Required By default, a port operates in noRestrictions mode.



On a port operating in either *macAddressElseUserLoginSecure* mode or *macAddressElseUserLoginSecureExt* mode, intrusion protection is triggered only after both MAC authentication and 802.1x authentication for the same frame fail.

Configuring Port Security Features

Configuring NTK Follow these steps to configure the NTK feature:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure the NTK feature	port-security ntk-mode { ntk-withbroadcasts ntk-withmulticasts ntkonly }	Required By default, NTK is disabled on a port and all frames are allowed to be sent.

Configuring Intrusion Protection Follow these steps to configure the intrusion protection feature:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	-
Configure the intrusion protection feature	port-security intrusion-mode { blockmac disableport disableport-temporarily }	Required By default, intrusion protection is disabled.
Return to system view	quit	-
Set the silence timeout during which a port remains disabled	port-security timer disableport <i>time-value</i>	Optional 20 seconds by default



If you configure the **port-security intrusion-mode** command with the **disableport-temporarily** keyword, you can use the **port-security timer disableport** command to set the silence timeout during which a port remains disabled.

Configuring Trapping Follow these steps to configure port security trapping:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable port security traps	port-security trap { addresslearned dot1xlogfailure dot1xlogoff dot1xlogon intrusion ralmlogfailure ralmlogoff ralmlogon }	Required By default, no port security trap is enabled.

Configuring Secure MAC Addresses

Secure MAC addresses are special MAC addresses. They never age out or get lost if saved before the device restarts. One secure MAC address can be added to only one port in the same VLAN. Thus, you can bind a MAC address to one port in the same VLAN.

Secure MAC addresses can be learned by a port working in autoLearn mode. You can also manually configure them through the command line interface (CLI) or management information base (MIB).

Configuration Prerequisites

- Enable port security
- Set the maximum number of secure MAC addresses allowed on the port
- Set the port security mode to autoLearn

Configuration Procedure

Follow these steps to configure a secure MAC address:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure a secure MAC address	In system view port-security mac-address security mac-address interface interface-type interface-number vlan vlan-id	Required Use either approach No secure MAC address is configured by default.
	In Ethernet port view interface interface-type interface-number port-security mac-address security mac-address vlan vlan-id	



The configured secure MAC addresses are saved in the configuration file and will not get lost when the port goes up or goes down. After you save the configuration file, the secure MAC address saved in the configuration file are maintained even after the device restarts.

Ignoring the Authorization Information from the Server

After an 802.1x user or MAC authenticated user passes RADIUS authentication, the RADIUS server delivers the authorization information to the device. You can configure a port to ignore the authorization information from the RADIUS server.

Follow these steps to configure a port to ignore the authorization information from the RADIUS server:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet port view	interface <i>interface-type</i> <i>interface-number</i>	-
Ignore the authorization information from the RADIUS server	port-security authorization ignore	Required By default, a port uses the authorization information from the RADIUS server.

Displaying and Maintaining Port Security

To do...	Use the command...	Remarks
Display port security configuration information, operation information, and statistics about one or more ports or all ports	display port-security [interface <i>interface-list</i>]	Available in any view
Display information about secure MAC addresses	display port-security mac-address security [interface <i>interface-type</i> <i>interface-number</i>] [vlan <i>vlan-id</i>] [count]	Available in any view
Display information about blocked MAC addresses	display port-security mac-address block [interface <i>interface-type</i> <i>interface-number</i>] [vlan <i>vlan-id</i>] [count]	Available in any view

Port Security Configuration Examples

Port Security Configuration for autoLearn Mode

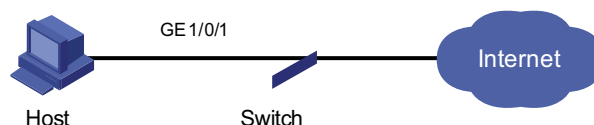
Network requirements

Restrict port GigabitEthernet 1/0/1 of the switch as follows:

- Allow up to 64 users to access the port without authentication and permit the port to learn and add the MAC addresses of the users as secure MAC addresses.
- After the number of secure MAC addresses reaches 64, the port stops learning MAC addresses. If any frame with an unknown MAC address arrives, intrusion protection is triggered and the port is disabled and stays silence for 30 seconds.

Network diagram

Figure 352 Network diagram for port security configuration for autoLearn mode



Configuration procedure

- 1 Configure port security

```
# Enable port security.
```

```

<Switch> system-view
[Switch] port-security enable

# Enable intrusion protection trap.

[Switch] port-security trap intrusion
[Switch] interface gigabitethernet 1/0/1

# Set the maximum number of secure MAC addresses allowed on the port to 64.

[Switch-GigabitEthernet1/0/1] port-security max-mac-count 64

# Set the port security mode to autoLearn.

[Switch-GigabitEthernet1/0/1] port-security port-mode autolearn

# Configure the port to be silent for 30 seconds after the intrusion protection
feature is triggered.

[Switch-GigabitEthernet1/0/1] port-security intrusion-mode disablepo
rt-temporarily
[Switch-GigabitEthernet1/0/1] quit
[Switch] port-security timer disableport 30

```

2 Verify the configuration

After completing the above configurations, you can use the following command to view the port security configuration information:

```

<Switch> display port-security interface gigabitethernet 1/0/1
Equipment port-security is enabled
Intrusion trap is enabled
Disableport Timeout: 30s
OUI value:

GigabitEthernet1/0/1 is link-up
Port mode is autoLearn
NeedToKnow mode is disabled
Intrusion Protection mode is DisablePortTemporarily
Max MAC address number is 64
Stored MAC address number is 0
Authorization is permitted

```

As shown in the output, the maximum number of secure MAC addresses allowed on the port is 64, the port security mode is autoLearn, the intrusion protection trap is enabled, and the intrusion protection action is to keep the port temporarily (DisablePortTemporarily) for 30 seconds.

You can also use the above command repeatedly to track the number of MAC addresses learned by the port, or use the **display this** command in Ethernet port view to display the secure MAC addresses learned, as shown below:

```

<Switch> system-view
[Switch] interface gigabitethernet 1/0/1
[Switch-GigabitEthernet1/0/1] display this
#
interface GigabitEthernet1/0/1

```

```

port-security max-mac-count 64
port-security port-mode autolearn
port-security mac-address security 0002-0000-0015 vlan 1
port-security mac-address security 0002-0000-0014 vlan 1
port-security mac-address security 0002-0000-0013 vlan 1
port-security mac-address security 0002-0000-0012 vlan 1
port-security mac-address security 0002-0000-0011 vlan 1
#

```

Issuing the **display port-security interface** command after the number of MAC addresses learned by the port reaches 64, you will see that the port security mode has changed to secure. When any frame with a new MAC address arrives, intrusion protection is triggered and you will see trap messages as follows:

```

#May 2 03:15:55:871 2000 Switch PORTSEC/1/VIOLATION:Traph3cSecureViolation
A intrusion occurs!
IfIndex: 9437207
Port: 9437207
MAC Addr: 0.2.0.0.0.21
VLAN ID: 1
IfAdminStatus: 1

```

In addition, you will see that the port security feature has disabled the port if you issue the following command:

```

[Switch-GigabitEthernet1/0/1] display interface gigabitethernet 1/0/1
GigabitEthernet1/0/1 current state: Port Security Disabled
IP Packet Frame Type: PKTFMT_ETHNT_2, Hardware Address: 000f-cb00-5558
Description: GigabitEthernet1/0/1 Interface
.....

```

The port should be re-enabled 30 seconds later.

```

[Switch-GigabitEthernet1/0/1] display interface gigabitethernet 1/0/1
GigabitEthernet1/0/1 current state: UP
IP Packet Frame Type: PKTFMT_ETHNT_2, Hardware Address: 000f-cb00-5558
Description: GigabitEthernet1/0/1 Interface
.....

```

Now, if you manually delete several secure MAC addresses, the port security mode of the port will be restored to autoLearn, and the port will be able to learn MAC addresses again.

Port Security Configuration for userLoginWithOUI Mode

Network requirements

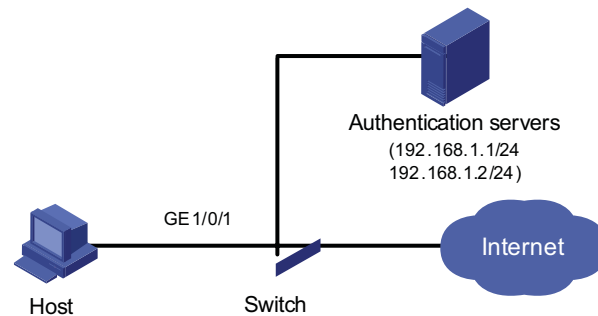
The client is connected to the switch through port GigabitEthernet 1/0/1. The switch authenticates the client by the RADIUS server. If the authentication succeeds, the client is authorized to access the Internet.

Restrict port GigabitEthernet 1/0/1 of the switch as follows:

- Allow only one 802.1x user to be authenticated.
- Allow up to 16 OUI values to be configured and allow one additional user whose MAC address has an OUI among the configured ones to access the port.

Network diagram

Figure 353 Network diagram for port security configuration for userLoginWithOUI mode



Configuration procedure



- The following configuration steps cover some AAA/RADIUS configuration commands. For details about the commands, refer to “Configuring AAA” on page 758 and “Configuring RADIUS” on page 765.
- Configurations on the host and RADIUS servers are omitted.

1 Configure the RADIUS protocol

Create a RADIUS scheme named radsun.

```
<Switch> system-view
[Switch] radius scheme radsun
```

Set the IP addresses of the primary authentication and accounting servers to 192.168.1.1 and 192.168.1.2 respectively.

```
[Switch-radius-radsun] primary authentication 192.168.1.1
[Switch-radius-radsun] primary accounting 192.168.1.2
```

Set the IP addresses of the secondary authentication and accounting servers to 192.168.1.2 and 192.168.1.1 respectively.

```
[Switch-radius-radsun] secondary authentication 192.168.1.2
[Switch-radius-radsun] secondary accounting 192.168.1.1
```

Set the encryption key for the switch to use when interacting with the authentication server to name.

```
[Switch-radius-radsun] key authentication name
```

Set the encryption key for the switch to use when interacting with the accounting server to money.

```
[Switch-radius-radsun] key accounting money
```

Set the RADIUS server response timeout time to five seconds and the maximum number of RADIUS packet retransmission attempts to 5.

```
[Switch-radius-radsun] timer response-timeout 5
[Switch-radius-radsun] retry 5
```


Set the interval at which the switch sends real-time accounting packets to the RADIUS server to 15 minutes.

```
[Switch-radius-radsun] timer realtime-accounting 15
```

Specify that the switch sends user names without domain names to the RADIUS server.

```
[Switch-radius-radsun] user-name-format without-domain
[Switch-radius-radsun] quit
```

Create an ISP domain named sun and enter its view.

```
[Switch] domain sun
```

Configure the ISP domain to use RADIUS scheme radsun as its default RADIUS scheme.

```
[Switch-isp-sun] authentication default radius-scheme radsun
```

Allow the ISP domain to accommodate up to 30 users.

```
[Switch-isp-sun] access-limit enable 30
[Switch-isp-sun] quit
```

2 Configure port security

Enable port security.

```
[Switch] port-security enable
```

Add five OUI values.

```
[Switch] port-security oui 1234-0100-1111 index 1
[Switch] port-security oui 1234-0200-1111 index 2
[Switch] port-security oui 1234-0300-1111 index 3
[Switch] port-security oui 1234-0400-1111 index 4
[Switch] port-security oui 1234-0500-1111 index 5
[Switch] interface gigabitethernet 1/0/1
```

Set the port security mode to userLoginWithOUI.

```
[Switch-GigabitEthernet1/0/1] port-security port-mode userlogin-withoui
```

3 Verify the configuration

After completing the above configurations, you can use the following command to view the configuration information of the RADIUS scheme named radsun:

```
<Switch> display radius scheme radsun
SchemeName = radsun
Index = 0
Primary Auth IP = 192.168.1.1      Port = 1812  State = active
Primary Acct IP = 192.168.1.2      Port = 1813  State = active
Second Auth IP = 192.168.1.2      Port = 1812  State = active
Second Acct IP = 192.168.1.1      Port = 1813  State = active
Auth Server Encryption Key = name
Acct Server Encryption Key = money
Accounting-On packet disable, send times = 5 , interval = 3s
Interval for timeout(second) = 5
```

```

Retransmission times for timeout                = 5
Interval for realtime accounting(minute)        = 15
Retransmission times of realtime-accounting packet = 5
Retransmission times of stop-accounting packet  = 500
Quiet-interval(min)                            = 5
Username format                                 = without-domain
Data flow unit                                  = Byte
Packet unit                                     = one

```

Use the following command to view the configuration information of the ISP domain named sun:

```

<Switch> display domain sun
  Domain = sun
  State = Active
  Access-limit = 30
  Accounting method = Required
  Default authentication scheme      : radius=radsun
  Default authorization scheme      : local
  Default accounting scheme         : local
  Domain User Template:
  Idle-cut = Disable
  Self-service = Disable

```

Use the following command to view the port security configuration information:

```

<Switch> display port-security interface gigabitethernet 1/0/1
Equipment port-security is enabled
Trap is disabled
Disableport Timeout: 20s
OUI value:
  Index is 1, OUI value is 123401
  Index is 2, OUI value is 123402
  Index is 3, OUI value is 123403
  Index is 4, OUI value is 123404
  Index is 5, OUI value is 123405

GigabitEthernet1/0/1 is link-up
  Port mode is userLoginWithOUI
  NeedToKnow mode is disabled
  Intrusion Protection mode is NoAction
  Max MAC address number is not configured
  Stored MAC address number is 0
  Authorization is permitted

```

After an 802.1x user gets online, you can see that the number of secure MAC addresses stored is 1. You can also use the following command to view information about 802.1x users:

```

<Switch> display dot1x interface gigabitethernet 1/0/1
Equipment 802.1X protocol is enabled
CHAP authentication is enabled

Configuration: Transmit Period   30 s, Handshake Period       15 s
                  Quiet Period   60 s, Quiet Period Timer is disabled
                  Supp Timeout    30 s, Server Timeout        100 s
                  The maximal retransmitting times           2

The maximum 802.1X user resource number is 1024 per slot
Total current used 802.1X resource number is 1

```

```
GigabitEthernet1/0/1 is link-up
  802.1X protocol is enabled
  Handshake is enabled
  The port is an authenticator
  Authentication Mode is Auto
  Port Control Type is Mac-based
  802.1X Multicast-trigger is enabled
  Guest VLAN: 0
  Max number of on-line users is 256

  EAPOL Packet: Tx 16331, Rx 102
  Sent EAP Request/Identity Packets : 16316
    EAP Request/Challenge Packets: 6
    EAP Success Packets: 4, Fail Packets: 5
  Received EAPOL Start Packets : 6
    EAPOL LogOff Packets: 2
    EAP Response/Identity Packets : 80
    EAP Response/Challenge Packets: 6
    Error Packets: 0
  1. Authenticated user : MAC address: 0002-0000-0011

  Controlled User(s) amount to 1
```

In addition, the port allows an additional user whose MAC address has an OUI among the specified OUIs to access the port. You can use the following command to view the related information:

```
<Switch> display mac-address interface gigabitEthernet 1/0/1
MAC ADDR      VLAN ID  STATE      PORT INDEX      AGING TIME(s)
1234-0300-0011  1       Learned    GigabitEthernet1/0/1  AGING

--- 1 mac address(es) found ---
```

Port Security Configuration for macAddressElseUserLogi nSecure Mode

Network requirements

The client is connected to the switch through GigabitEthernet 1/0/1. The switch authenticates the client by the RADIUS server. If the authentication succeeds, the client is authorized to access the Internet.

Restrict port GigabitEthernet 1/0/1 of the switch as follows:

- Allow more than one MAC authenticated user to log on.
- For 802.1x users, perform MAC authentication first and then, if MAC authentication fails, 802.1x authentication. Allow only one 802.1x user to log on.
- For MAC-based authentication, allow usernames and passwords in self-defined formats. Set the total number of MAC authenticated users and 802.1x-authenticated users to 64.
- Enable NTK to prevent frames from being sent to unknown MAC addresses.

Network diagram

See Figure 353.

Configuration procedure

Configurations on the host and RADIUS servers are omitted.

1 Configure the RADIUS protocol

The required RADIUS authentication/accounting configurations are the same as those in “Port Security Configuration for userLoginWithOUI Mode” on page 1171.

2 Configure port security

Enable port security.

```
<Switch> system-view
[Switch] port-security enable
```

Configure a MAC authentication user, setting the user name and password to aaa and 123456 respectively.

```
[Switch] mac-authentication user-name-format fixed account aaa password simple 123456
[Switch] interface gigabitethernet 1/0/1
```

Set the maximum number of secure MAC addresses allowed on the port to 64.

```
[Switch-GigabitEthernet1/0/1] port-security max-mac-count 64
```

Set the port security mode to macAddressElseUserLoginSecure.

```
[Switch-GigabitEthernet1/0/1] port-security port-mode mac-else-userlogin-secure
```

Set the NTK mode of the port to ntkonly.

```
[Switch-GigabitEthernet1/0/1] port-security ntk-mode ntkonly
```

3 Verify the configuration

After completing the above configurations, you can use the following command to view the port security configuration information:

```
<Switch> display port-security interface gigabitethernet 1/0/1
Equipment port-security is enabled
Trap is disabled
Disableport Timeout: 20s
OUI value:
```

```
GigabitEthernet1/0/1 is link-up
Port mode is macAddressElseUserLoginSecure
NeedToKnow mode is NeedToKnowOnly
Intrusion Protection mode is NoAction
Max MAC address number is 64
Stored MAC address number is 0
Authorization is permitted
```

Use the following command to view MAC authentication information:

```
<Switch> display mac-authentication interface gigabitethernet 1/0/1
MAC address authentication is enabled.
User name format is fixed account
Fixed username:aaa
Fixed password:123456
    Offline detect period is 300s
    Quiet period is 60s
    Server response timeout value is 100s
    The max allowed user number is 1024 per slot
    Current user number amounts to 0
    Current domain: not configured, use default domain
```

Silent MAC User info:

MAC Addr	From Port	Port Index
GigabitEthernet1/0/1 is link-up		
MAC address authentication is enabled		
Authenticate success: 3, failed: 1		
Current online user number is 3		
MAC Addr	Authenticate State	Auth Index
1234-0300-0011	MAC_AUTHENTICATOR_SUCCESS	13
1234-0300-0012	MAC_AUTHENTICATOR_SUCCESS	14
1234-0300-0013	MAC_AUTHENTICATOR_SUCCESS	15

Use the following command to view 802.1x authentication information:

```
<Switch> display dot1x interface gigabitethernet 1/0/1
Equipment 802.1X protocol is enabled
CHAP authentication is enabled

Configuration: Transmit Period 30 s, Handshake Period 15 s
                Quiet Period 60 s, Quiet Period Timer is disabled
                Supp Timeout 30 s, Server Timeout 100 s
                The maximal retransmitting times 2

The maximum 802.1X user resource number is 1024 per slot
Total current used 802.1X resource number is 1

GigabitEthernet1/0/1 is link-up
802.1X protocol is enabled
Handshake is enabled
The port is an authenticator
Authentication Mode is Auto
Port Control Type is Mac-based
802.1X Multicast-trigger is enabled
Guest VLAN: 0
Max number of on-line users is 256

EAPOL Packet: Tx 16331, Rx 102
Sent EAP Request/Identity Packets : 16316
    EAP Request/Challenge Packets: 6
    EAP Success Packets: 4, Fail Packets: 5
Received EAPOL Start Packets : 6
    EAPOL LogOff Packets: 2
    EAP Response/Identity Packets : 80
    EAP Response/Challenge Packets: 6
    Error Packets: 0
1. Authenticated user : MAC address: 0002-0000-0011

Controlled User(s) amount to 1
```

In addition, since NTK is enabled, frames with unknown destination MAC addresses, multicast addresses, and broadcast addresses should be discarded.

Troubleshooting Port Security

Cannot Set the Port Security Mode

Symptom

Cannot set the port security mode.

```
[Switch-GigabitEthernet1/0/1] port-security port-mode autolearn
```

Error:When we change port-mode, we should first change it to noRestrictions, then change it to the other.

Analysis

For a port working in a port security mode other than noRestrictions, you cannot change the port security mode by using the **port-security port-mode** command directly.

Solution

Set the port security mode to noRestrictions first.

```
[Switch-GigabitEthernet1/0/1] undo port-security port-mode
[Switch-GigabitEthernet1/0/1] port-security port-mode autolearn
```

Cannot Configure Secure MAC Addresses

Symptom

Cannot configure secure MAC addresses.

```
[Switch-GigabitEthernet1/0/1] port-security mac-address security 1-1-2 vlan 1
Error:Can not operate security MAC address for current port mode is not autoLearn!
```

Analysis

No secure MAC address can be configured on a port operating in a port security mode other than autoLearn.

Solution

Set the port security mode to autoLearn.

```
[Switch-GigabitEthernet1/0/1] undo port-security port-mode
[Switch-GigabitEthernet1/0/1] port-security max-mac-count 64
[Switch-GigabitEthernet1/0/1] port-security port-mode autolearn
[Switch-GigabitEthernet1/0/1] port-security mac-address security 1-1-2 vlan 1
```

Cannot Change Port Security Mode When a User Is Online

Symptom

Port security mode cannot be changed when an 802.1x-authenticated or MAC authenticated user is online.

```
[Switch-GigabitEthernet1/0/1] undo port-security port-mode
```

Error:Cannot configure port-security for there is 802.1X user(s) on line on port GigabitEthernet1/0/1.

Analysis

Changing port security mode is not allowed when an 802.1x-authenticated or MAC authenticated user is online.

Solution

Use the **cut** command to forcibly disconnect the user from the port before changing the port security mode.

```
[Switch-GigabitEthernet1/0/1] cut connection interface gigabitethernet 1/0/1  
[Switch-GigabitEthernet1/0/1] undo port-security port-mode
```


When configuring LLDP, go to these sections for information you are interested in:

- “Introduction to LLDP” on page 1181
- “LLDP Configuration Tasks List” on page 1184
- “Performing Basic LLDP Configuration” on page 1184
- “Configuring LLDP Trap” on page 1188
- “Displaying and Maintaining LLDP” on page 1188
- “LLDP Configuration Example” on page 1189

Introduction to LLDP

Overview Link Layer Discovery Protocol (LLDP) operates on data link layer. It stores and maintains the information about the local device and the devices directly connected to it for network administrators to manage networks through NMS (network management systems). In LLDP, device information is encapsulated in LLDPDUs in the form of TLV (meaning type, length, and value) triplets and is exchanged between directly connected devices. Information in LLDPDUs received is restored in standard MIB (management information base).

LLDP Fundamental **LLDP operating mode**

LLDP can operate in one of the following modes.

- TxRx mode. A port in this mode sends and receives LLDPDUs.
- Tx mode. A port in this mode only sends LLDPDUs.
- Rx mode. A port in this mode only receives LLDPDUs.
- Disable mode. A port in this mode does not send or receive LLDPDUs.

LLDP is initialized when an LLDP-enabled port changes to operate in another LLDP operating mode. To prevent LLDP from being initialized too frequently, LLDP undergoes a period before being initialized on an LLDP-enabled port when the port changes to operate in another LLDP operating mode. The period is known as initialization delay, which is determined by the re-initialization delay timer.

Sending LLDPDUs

An LLDP-enabled device operating in the TxRx mode or Tx mode sends LLDPDUs to its directly connected devices periodically. It also sends LLDPDUs when the local configuration changes to inform the neighboring devices of the change timely. In any of the two cases, an interval exists between two successive operations of sending LLDPDUs. This prevents the network from being overwhelmed by LLDPDUs even if the LLDP operating mode changes frequently.

To enable the neighboring devices to be informed of the existence of a device or an LLDP operating mode change (from the disable mode to TxRx mode, or from the Rx mode to Tx mode) timely, a device can invoke the fast sending mechanism. In this case, the interval to send LLDPDUs changes to one second. After the device sends specific number of LLDPDUs, the interval restores to the normal. (A neighbor is discovered when a device receives an LLDPDU and no information about the sender is locally available.)

Receiving LLDPDUs

An LLDP-enabled device operating in the TxRx mode or Rx mode checks the TLVs carried in the LLDPDUs it receives and saves the valid neighboring information. An LLDPDU also carries a TTL (time to live) setting with it. The information about a neighboring device maintained locally ages out when the corresponding TTL expires.

The TTL of the information about a neighboring device is determined by the following expression:

TTL multiplier × LLDPDU sending interval.

You can set the TTL by configuring the TTL multiplier. Note that the TTL can be up to 65535 seconds. TTLs longer than it will be rounded off to 65535 seconds.

TLV Types TLVs encapsulated in LLDPDUs fall into these categories: basic TLV, organization defined TLV, and MED (media endpoint discovery) related TLV. Basic TLVs are the base of device management. Organization specific TLVs and MED related TLVs are used for enhanced device management. They are defined in standards or by organizations and are optional to LLDPDUs.

Basic LLDP TLVs

Table 91 lists the basic LLDP TLV types that are currently in use.

Table 91 Basic LLDP TLVs

Type	Description	Remarks
End of LLDPDU TLV	Marks the end of an LLDPDU.	Required for LLDP
Chassis ID TLV	Carries the bridge MAC address of the sender	
Port ID TLV	Carries the sending port. For devices that do not send MED TLVs, port ID TLVs carry sending port name. For devices that send MED TLVs, port ID TLVs carry the MAC addresses of the sending ports or bridge MAC addresses (if the MAC addresses of the sending ports are unavailable).	
Time To Live TLV	Carries the TTL of device information	

Table 91 Basic LLDP TLVs

Type	Description	Remarks
Port Description TLV	Carries Ethernet port description	Optional to LLDP
System Name TLV	Carries device name	
System Description TLV	Carries system description	
System Capabilities TLV	Carries information about system capabilities	
Management Address TLV	Carries the management address, the corresponding port number, and OID (object identifier). If the management address is not configured, it is the IP address of the interface of the VLAN with the least VLAN ID among those permitted on the port. If the IP address of the VLAN interface is not configured, IP address 127.0.0.1 is used as the management address.	

Organization defined LLDP TLVs

- LLDP TLVs defined in IEEE802.1 include the following:
 - Port VLAN ID TLV, which carries port VLAN ID.
 - Port and protocol VLAN ID TLV, which carries port protocol VLAN ID.
 - VLAN name TLV, which carries port VLAN name.
 - Protocol identity TLV, which carries types of the supported protocols.



Currently, protocol identity TLVs can only be received on 3Com devices.

- IEEE 802.3 defined LLDP TLVs include the following:
 - MAC/PHY configuration/status TLV, which carries port configuration, such as port speed, duplex state, whether port speed auto-negotiation is supported, the state of auto-negotiation, current speed, and current duplex state.
 - Power via MDI TLV, which carries information about power supply capabilities.
 - Link aggregation TLV, which carries the capability and state of link aggregation.
 - Maximum frame size TLV, which carries the maximum frame size supported, namely, MTU (maximum transmission unit).

MED related LLDP TLVs

- LLDP-MED capabilities TLV, which carries the MED type of the current device and the types of the LLDP MED TLVs that can be encapsulated in LLDPDUs.
- Network policy TLV, which carries port VLAN ID, supported applications (such as voice and video services), application priority, and the policy adopted.
- Extended power-via-MDI TLV, which carries the information about the power supply capability of the current device.
- Hardware revision TLV, which carries the hardware version of an MED device.
- Firmware revision TLV, which carries the firmware version of an MED device.
- Software revision TLV, which carries the software version of an MED device.
- Serial number TLV, which carries the serial number of an MED device.

- Manufacturer name TLV, which carries the manufacturer name of an MED device.
- Model name TLV, which carries the model of an MED device.
- Asset ID TLV, which carries the asset ID of an MED device. Asset ID is used for directory management and asset tracking.
- Location identification TLV, which carries the location identification of a device. Location identification can be used in location-based applications.



For detailed information about LLDP TLV, refer to IEEE 802.1AB-2005 and ANSI/TIA-1057.

Protocols and Standards

- IEEE 802.1AB-2005, Station and Media Access Control Connectivity Discovery
- ANSI/TIA-1057, Link Layer Discovery Protocol for Media Endpoint Devices

LLDP Configuration Tasks List

Complete these tasks to configure LLDP:

Task	Remarks	
Basic LLDP configuration	"Enabling LLDP" on page 1184	Required
	"Setting LLDP Operating Mode" on page 1185	Optional
	"Configuring LLDPDU TLVs" on page 1185	Optional
	"Enable LLDP Polling" on page 1186	Optional
	"Configuring the Parameters Concerning LLDPDU Sending" on page 1186	Optional
"Configuring LLDP Trap" on page 1188	Optional	

Performing Basic LLDP Configuration

Enabling LLDP

Follow these steps to enable LLDP:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable LLDP globally	lldp enable	Required The default global state of LLDP varies with device models.
Enter Ethernet interface view/port group view	Enter Ethernet interface view Enter port group view	Either of the two is required. Configuration performed in Ethernet interface view applies to the current port only; configuration performed in port group view applies to all the ports in the corresponding port group.
Enable LLDP	lldp enable	Optional By default, LLDP is enabled on a port.



To make LLDP take effect, you need to enable it both globally and on the related ports.

Setting LLDP Operating Mode

Follow these steps to set LLDP operating mode:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Set the initialization delay period	lldp timer reinit-delay <i>value</i>	Optional 2 seconds by default.
Enter Ethernet interface view/port group view	Enter Ethernet interface view interface <i>interface-type interface-number</i> Enter port group view port-group { aggregation <i>agg-id</i> manual <i>port-group-name</i> }	Either of the two is required. Configuration performed in Ethernet interface view applies to the current port only; configuration performed in port group view applies to all the ports in the corresponding port group.
Set the LLDP operating mode	lldp admin-status { disable rx tx txrx }	Optional TxRx by default.

Configuring LLDPDU TLVs

Follow these steps to configure LLDPDU TLVs:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Set the TTL multiplier	lldp hold-multiplier <i>value</i>	Optional 4 by default.
Enter Ethernet interface view/port group view	Enter Ethernet interface view interface <i>interface-type interface-number</i> Enter port group view port-group { aggregation <i>agg-id</i> manual <i>port-group-name</i> }	Either of the two is required. Configuration performed in Ethernet interface view applies to the current port only; configuration performed in port group view applies to all the ports in the corresponding port group.
Enable LLDP TLV sending for specific types of LLDP TLVs	lldp tlv-enable { basic-tlv { all port-description system-capability system-description system-name } dot1-tlv { all port-vlan-id protocol-vlan-id [<i>vlan-id</i>] vlan-name [<i>vlan-id</i>] } dot3-tlv { all link-aggregation mac-physic max-frame-size power } med-tlv { all capability location-id { civic-address <i>device-type country-code</i> { <i>ca-type ca-value</i> } &<1-10> elin-address <i>Tel-Number</i> } network-policy power-over-ethernet inventory } }	Optional By default, all types of LLDP TLVs except location identification TLV are sent.

To do...	Use the command...	Remarks
Specify the management address and specify to send the management address through LLDPDUs	lldp management-address-tlv [<i>ip-address</i>]	Optional By default, the management address is sent through LLDPDUs, and the management address is the IP address of the interface of the VLAN with the least VLAN ID among those permitted on the port. If the IP address of the VLAN interface is not configured, IP address 127.0.0.1 is used as the management address. Refer to "VLAN Configuration" on page 83.



- *To enable MED related LLDP TLV sending, you need to enable LLDP-MED capabilities TLV sending first. Conversely, to disable LLDP-MED capabilities TLV sending, you need to disable the sending of other MED related LLDP TLVs.*
- *To disable MAC/PHY configuration/status TLV sending, you need to disable LLDP-MED capabilities TLV sending first.*
- *When executing the **lldp tlv-enable** command, specifying the **all** keyword for basic LLDP TLVs and organization defined LLDP TLVs (including IEEE 802.1 defined LLDP TLVs and IEEE 802.3 defined LLDP TLVs) enables sending of all the corresponding LLDP TLVs. For MED related LLDP TLVs, the **all** keyword enables sending of all the MED related LLDP TLVs except location identification TLVs.*
- *Enabling sending of LLDP-MED capabilities TLVs also enables sending of MAC/PHY configuration/status TLVs.*

Enable LLDP Polling

With LLDP polling enabled, a device checks for the local configuration changes periodically. Upon detecting a configuration change, the device sends LLDPDUs to inform the neighboring devices of the change.

Follow these steps to enable LLDP polling:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet interface view/port group view	Enter Ethernet interface view interface <i>interface-type interface-number</i> Enter port group view port-group { aggregation <i>agg-id</i> manual <i>port-group-name</i> }	Either of the two is required. Configuration performed in Ethernet interface view applies to the current port only; configuration performed in port group view applies to all the ports in the corresponding port group.
Enable LLDP polling and set the polling interval	lldp check-change-interval <i>value</i>	Optional Disabled by default

Configuring the Parameters Concerning LLDPDU Sending

Configuring time-related parameters

Follow these steps to set time-related parameters:

To do...	Use the command...	Remarks
Enter system view	System-view	-
Set the interval to send LLDPDUs	lldp timer tx-interval <i>value</i>	Optional 30 seconds by default
Set the delay period to send LLDPDUs	lldp timer tx-delay <i>value</i>	Optional 2 seconds by default



CAUTION: To enable local device information to be updated on neighboring devices before being aged out, make sure the interval to send LLDPDUs is shorter than the TTL of the local device information.

Setting the number of the LLDPDUs to be sent when a new neighboring device is detected

Follow these steps to set the number of the LLDPDUs to be sent when a new neighboring device is detected

To do...	Use the command...	Remarks
Enter system view	system-view	-
Set the number of the LLDPDUs to be sent successively when a new neighboring device is detected	lldp fast-count <i>value</i>	Optional 3 by default

Configuring the Encapsulation Format for LLDPDUs

The Switch 4800G can encapsulate LLDPDUs in Ethernet II or SNAP frames. You can configure either encapsulation at the CLI for interoperability with the remote device.

- With Ethernet II encapsulation configured, an LLDP port sends LLDPDUs in Ethernet II encapsulation and processes only Ethernet II encapsulated incoming LLDPDUs.
- With SNAP encapsulation configured, an LLDP port sends LLDPDUs in SNAP encapsulation and processes only SNAP encapsulated incoming LLDPDUs.

Follow these steps to configure the encapsulation format for LLDPDUs:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet interface view or port group view	Enter Ethernet interface view Enter port group view	Either of the two is required.
	port-group { aggregation <i>agg-id</i> manual <i>port-group-name</i> }	Configuration performed in Ethernet interface view applies to the current port only; configuration performed in port group view applies to all the ports in the corresponding port group.

To do...	Use the command...	Remarks
Configure the encapsulation format for LLDPDUs as SNAP	lldp encapsulation snap	Optional
Configure the encapsulation format for LLDPDUs as Ethernet II	undo lldp encapsulation [snap]	Ethernet II encapsulation format applies by default.



The configuration does not apply to LLDP-CDP packets, which use only SNAP encapsulation.

Configuring LLDP Trap

LLDP trap is used to notify NMS of the events such as new neighboring devices detected and link malfunctions.

LLDP traps are sent periodically and you can set the interval to send LLDP traps. In response to topology changes detected, a device sends LLDP traps according to the interval configured to inform the neighboring devices of the changes.

Follow these steps to configure LLDP trap:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter Ethernet interface view/port group view	interface <i>interface-type interface-number</i> port-group { aggregation agg-id manual port-group-name }	Either of the two is required. Configuration performed in Ethernet interface view applies to the current port only; configuration performed in port group view applies to all the ports in the corresponding port group.
Enable LLDP trap sending	lldp notification remote-change enable	Required Disabled by default
Quit to system view	quit	-
Set the interval to send LLDP traps	lldp timer notification-interval <i>value</i>	Optional 5 seconds by default

Displaying and Maintaining LLDP

To do...	Use the command...	Remarks
Display the global LLDP information or the information contained in the LLDP TLVs to be sent through a port	display lldp local-information [global interface <i>interface-type interface-number</i>]	Available in any view
Display the information contained in the LLDP TLVs received through a port	display lldp neighbor-information [interface <i>interface-type interface-number</i>] [brief]	Available in any view
Display LLDP statistics	display lldp statistics [global interface <i>interface-type interface-number</i>]	Available in any view

To do...	Use the command...	Remarks
Display LLDP status of a port	display lldp status [interface <i>interface-type</i> <i>interface-number</i>]	Available in any view
Display the types of the LLDP TLVs that are currently sent	display lldp tlv-config [interface <i>interface-type</i> <i>interface-number</i>]	Available in any view

LLDP Configuration Example

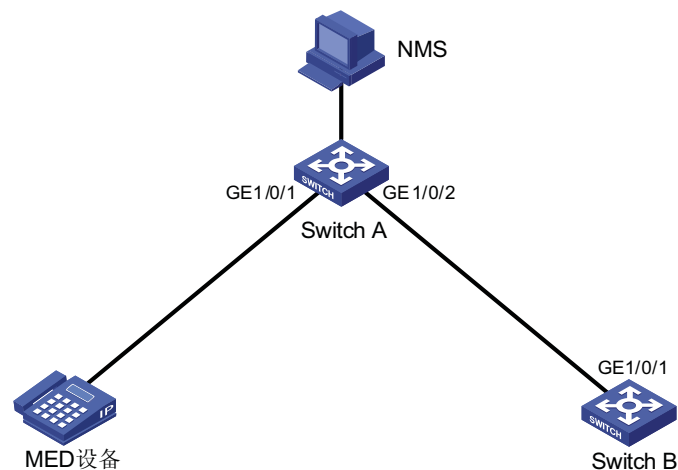
LLDP Configuration Example

Network requirements

- The NMS and Switch A are located in the same Ethernet. An MED device and Switch B are connected to GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 of Switch A.
- Enable LLDP on the ports of Switch A and Switch B to monitor the link between Switch A and Switch B and the link between Switch A and the MED device on the NMS.

Network diagram

Figure 354 Network diagram for LLDP configuration



Configuration procedure

- 1 Configure Switch A.

Enter system view.

```
<SwitchA> system-view
```

Enable LLDP globally.

```
[SwitchA] lldp enable
```

Enable LLDP on GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2, setting the LLDP operating mode to Rx.

```
[SwitchA] interface GigabitEthernet1/0/1
[SwitchA-GigabitEthernet1/0/1] lldp enable
[SwitchA-GigabitEthernet1/0/1] lldp admin-status rx
[SwitchA-GigabitEthernet1/0/1] quit
[SwitchA] interface GigabitEthernet1/0/2
[SwitchA-GigabitEthernet1/0/2] lldp enable
[SwitchA-GigabitEthernet1/0/2] lldp admin-status rx
[SwitchA-GigabitEthernet1/0/2] quit
```

2 Configure Switch B.

Enter system view.

```
<SwitchB> system-view
```

Enable LLDP globally.

```
[SwitchB] lldp enable
```

Enable LLDP on GigabitEthernet1/0/1, setting the LLDP operating mode to Tx.

```
[SwitchB] interface GigabitEthernet1/0/1
[SwitchB-GigabitEthernet1/0/1] lldp enable
[SwitchB-GigabitEthernet1/0/1] lldp admin-status tx
```

3 Verify the configuration.

Display the global LLDP status and port LLDP status on Switch A.

```
<SwitchA> display lldp status
Global status of LLDP : Enable
The current number of neighbors : 2
Neighbor information last changed time : 0 days, 0 hours, 4 minutes, 40 seconds
Transmit interval           : 30s
Hold multiplier             : 4
Reinit delay                : 2s
Transmit delay              : 2s
Trap interval               : 5s
Fast start times            : 3

Port 0 [GigabitEthernet1/0/1] :
Port status of LLDP         : Enable
Admin status                 : Rx_Only
Trap flag                    : No
Roll time                    : 0s

Number of neighbors         : 1
Number of MED neighbors     : 1
Number of sent optional TLV : 0
Number of received unknown TLV : 0

Port 1 [GigabitEthernet1/0/2] :
Port status of LLDP         : Enable
Admin status                 : Rx_Only
Trap flag                    : No
Roll time                    : 0s

Number of neighbors         : 1
Number of MED neighbors     : 0
Number of sent optional TLV : 0
Number of received unknown TLV : 3
```

Tear down the link between Switch A and Switch B and then display the global LLDP status and port LLDP status on Switch A.

```
<SwitchA> display lldp status
Global status of LLDP : Enable
The current number of neighbors : 1
Neighbor information last changed time : 0 days, 0 hours, 5 minutes, 20 seconds
Transmit interval          : 30s
Hold multiplier            : 4
Reinit delay               : 2s
Transmit delay             : 2s
Trap interval              : 5s
Fast start times           : 3

Port 0 [GigabitEthernet1/0/1] :
Port status of LLDP        : Enable
Admin status                : Rx_Only
Trap flag                   : No
Roll time                   : 0s

Number of neighbors        : 1
Number of MED neighbors    : 1
Number of sent optional TLV : 0
Number of received unknown TLV : 5

Port 1 [GigabitEthernet1/0/2] :
Port status of LLDP        : Enable
Admin status                : Rx_Only
Trap flag                   : No
Roll time                   : 0s

Number of neighbors        : 0
Number of MED neighbors    : 0
Number of sent optional TLV : 0
Number of received unknown TLV : 0
```


When configuring PoE, go to these sections for information you are interested in:

- "PoE Overview" on page 1193
- "PoE Configuration Task List" on page 1194
- "Configuring the PoE Interface" on page 1194
- "Configuring PD Power Management" on page 1196
- "Configuring a Power Alarm Threshold for the PSE" on page 1197
- "Upgrading PSE Processing Software Online" on page 1197
- "Configuring a PD Disconnection Detection Mode" on page 1198
- "Enabling the PSE to Detect Nonstandard PDs" on page 1198
- "Displaying and Maintaining PoE" on page 1199
- "PoE Configuration Example" on page 1199
- "Troubleshooting PoE" on page 1200

PoE Overview

Introduction to PoE Power over Ethernet (PoE) means that power sourcing equipment (PSE) supplies power to powered devices (PD) such as IP telephone, wireless LAN access point, and web camera from Ethernet interfaces through twisted pair cables.

Advantages

- **Reliable:** Power is supplied in a centralized way so that it is very convenient to provide a backup power supply.
- **Easy to connect:** A network terminal requires only one Ethernet cable, but no external power supply.
- **Standard:** In compliance with IEEE 802.3af, and a globally uniform power interface is adopted.
- **Promising:** It can be applied to IP telephones, wireless LAN access points, portable chargers, module readers, web cameras, and data collectors.

Composition

A PoE system consists of PoE power, PSE, and PD.

- PoE power

The whole PoE system is powered by the PoE power, which includes external PoE power and internal PoE power.

- PSE

PSE is a module or subcard. PSE manages its own PoE interfaces independently. PSE examines the Ethernet cables connected to PoE interfaces, searches for the devices, classifies them, and supplies power to them. When detecting that a PD is unplugged, the PSE stops supplying power to the PD.

An Ethernet interface with the PoE capability is called PoE interface. Currently, a PoE interface can be an FE or GE interface.

- PD

A PD is a device accepting power from the PSE. There are standard PDs and nonstandard PDs. A standard PD refers to the one that complies with IEEE 802.3af. The PD that is being powered by the PSE can be connected to other power supply units for redundancy backup.

Protocol Specification The protocol specification related to PoE is IEEE 802.3af.

PoE Configuration Task List

Complete these tasks to configure PoE:

Task	Remarks
"Configuring the PoE Interface" on page 1194	Required
"Configuring PD Power Management" on page 1196	Optional
"Configuring a Power Alarm Threshold for the PSE" on page 1197	Optional
"Upgrading PSE Processing Software Online" on page 1197	Optional
"Configuring a PD Disconnection Detection Mode" on page 1198	Optional
"Enabling the PSE to Detect Nonstandard PDs" on page 1198	Optional

Configuring the PoE Interface

You can configure a PoE interface in either of the following two ways:

- Adopting the command line.
- Configuring a PoE configuration file and applying the file to the specified PoE interface(s).

Usually, you can adopt the command line to configure a single PoE interface, and adopt a PoE configuration file to configure multiple PoE interfaces at the same time.



CAUTION: You can adopt either mode to configure, modify, or delete a PoE configuration parameter under the same PoE interface.

The PSE supplies power for a PoE interface in the following two modes:

- For a device with only signal cables, power is supplied over signal cables.
- For a device with spare cables and signal cables, power can be supplied over spare cables or signal cables.



the Switch 4800G do not support power over spare cables. In this case, if the PD only supports power over spare cables, you have to change the order of the lines in the twisted pair cable to supply power to the PD.

Configuring a PoE Interface through the Command Line

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PoE interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable PoE	poe enable	Required Disabled by default.
Configure the maximum power for the PoE interface	poe max-power <i>max-power</i>	Optional 15,400 milliwatts by default.
Configure the PoE mode for the PoE interface	poe mode signal	Optional signal (power over signal cables) by default.
Configure a description for the PD connected to the PoE interface	poe pd-description <i>string</i>	Optional By default, no description for the PD connected to the PoE interface is available.

Configuring PoE Interfaces Through a PoE Configuration File

A PoE configuration file is used to configure at the same time multiple PoE interfaces with the same attributes to simplify operations. This configuration method is a supplement to the command line configuration.

Commands in a PoE configuration file are called configurations.

Follow these steps to configure PoE interfaces through a PoE configuration file:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a PoE configuration file and enter PoE configuration file view	poe-profile <i>profile-name</i> [<i>index</i>]	Required
Enable PoE for the PoE interface	poe enable	Required Disabled by default.
Configure the maximum power for the PoE interface	poe max-power <i>max-power</i>	Optional 15,400 milliwatts by default.
Configure the PoE mode for the PoE interface	poe mode signal	Optional signal (power over signal cables) by default.
Return to system view	quit	-

To do...		Use the command...	Remarks
Apply the PoE configuration file to the PoE interface(s)	Apply the PoE configuration file to one or more PoE interfaces	apply poe-profile { index <i>index</i> name <i>profile-name</i> } interface <i>interface-range</i>	Use either approach
	Apply the PoE configuration file to the current PoE interface in PoE interface view	interface <i>interface-type</i> <i>interface-number</i> apply poe-profile { index <i>index</i> name <i>profile-name</i> }	

**CAUTION:**

- After a PoE configuration file is applied to a PoE interface, other PoE configuration files can not take effect on this PoE interface.
- If a PoE configuration file is already applied to a PoE interface, you must execute the **undo apply poe-profile** command to remove the application to the interface before deleting or modifying the PoE configuration file.
- If you have configured a PoE interface through the command line, you cannot configure it through a PoE configuration file again. If you want to reconfigure the interface through a PoE configuration file, you must first remove the command line configuration on the PoE interface.
- You must use the same mode (command line or PoE configuration file) to configure the **poe max-power** *max-power* and **poe priority** { **critical** | **high** | **low** } commands.

Configuring PD Power Management

The power priority of a PD depends on the priority of the PoE interface. The priority levels of PoE interfaces include critical, high and low in descending order. Power supply to a PD is subject to PD power management policies.

All PSEs implement the same PD power management policies. When the PSE supplies power to a PD,

- By default, no power will be supplied to a new PD if the PSE power is overloaded.
- Under the control of a priority policy, the PD with a lower priority is first powered off to guarantee the power supply to the new PD with a higher priority when the PSE power is overloaded.



If the sudden increase of the power of the PD results in PSE power overload, power supply to the PD on the PoE interface with a lower priority will be stopped.

If the guaranteed remaining PSE power (power allocated to the critical PoE interface subtracted from maximum PSE power, regardless of whether PoE is enabled for the PoE interface) is lower than the maximum power of the PoE interface, you will fail to set the priority of the PoE interface to **critical**. Otherwise, you can succeed in setting the priority to **critical**, and this PoE interface will preempt the power of other PoE interfaces with a lower priority level. In the latter case, the PoE interfaces whose power is preempted will be powered off, but their configurations will remain unchanged. When you change the priority of a PoE

interface from critical to a lower level, the PDs connecting to other PoE interfaces will have an opportunity of being powered.

Configuration prerequisites

Enable PoE for PoE interfaces.

Configuration procedure

Follow these steps to configure PD power management:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the power priority for a PoE interface	Configure the power priority for the PoE interface in PoE interface view interface <i>interface-type</i> <i>interface-number</i> poe priority { critical high low }	Use either command. By default, the power priority of a PoE interface is low .
	Configure the power priority for the PoE interface in PoE configuration file view poe-profile <i>profile-name</i> [<i>index</i>] poe priority { critical high low }	
Configure a PD power management priority policy	poe pd-policy priority	Optional By default, no PD power management priority policy is configured.

Configuring a Power Alarm Threshold for the PSE

When the current power utilization of the PSE is above or below the alarm threshold for the first time, the system will send a Trap message.

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure a power alarm threshold for the PSE	poe utilization-threshold <i>utilization-threshold-value</i>	Optional 80% by default.

Upgrading PSE Processing Software Online

You can upgrade the PSE processing software online in either of the following two modes:

- refresh mode

This mode enables you to update the PSE processing software without deleting it. Normally, you can upgrade the PSE processing software in the refresh mode through the command line.

- full mode

This mode deletes the PSE processing software and reloads it. When the PSE processing software is damaged (in this case, you can execute none of PoE commands successfully), you can upgrade the PSE software processing software in full mode to restore the PSE function.

Online PSE processing software upgrade may be unexpectedly interrupted (for example, an error results in device reboot). If you fail to upgrade the PSE processing software in full mode after reboot, you can power off the device and restart it before upgrading it again. After upgrade, restart the device manually to make the original PoE configurations take effect.

Follow these steps to upgrade the PSE processing software online:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Upgrade the PSE processing software online	poe update { full refresh } filename	Optional

Configuring a PD Disconnection Detection Mode

To detect the PD connection with PSE, PoE provides two detection modes: AC detection and DC detection. The AC detection mode is energy saving relative to the DC detection mode.

Follow these steps to configure a PD disconnection detection mode:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure a PD disconnection detection mode	poe disconnect { ac dc }	Optional The default PD disconnection detection mode is AC.



CAUTION: *If you adjust the PD disconnection detection mode when the device is running, the connected PDs will be powered off. Therefore, be cautious to do so.*

Enabling the PSE to Detect Nonstandard PDs

There are standard PDs and nonstandard PDs. Usually, the PSE can detect only standard PDs and supply power to them. The PSE can detect nonstandard PDs and supply power to them only after the PSE is enabled to detect nonstandard PDs.

Follow these steps to enable the PSE to detect nonstandard PDs:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the PSE to detect nonstandard PDs	poe legacy enable	Optional Disabled by default.

Displaying and Maintaining PoE

To do...	Use the command...	Remarks
Display the mapping between ID, module, and slot of all PSEs.	display poe device	Available in any view
Display the power state and information of the specified PoE interface	display poe interface [<i>interface-type interface-number</i>]	
Display the power information of a PoE interface(s)	display poe interface power [<i>interface-type interface-number</i>]	
Display the information of PSE	display poe pse	
Display all information of the configurations and applications of the PoE configuration file	display poe-profile [index <i>index</i> name <i>profile-name</i>]	
Display all information of the configurations and applications of the PoE configuration file applied to the specified PoE interface	display poe-profile interface <i>interface-type interface-number</i>	

PoE Configuration Example

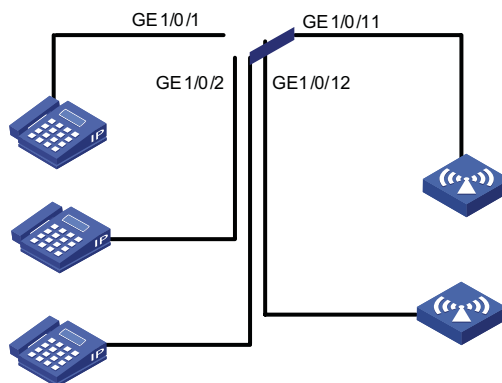
Network requirements

The device provides power supply for PDs through PoE interfaces.

- GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 are connected to IP telephones.
- GigabitEthernet 1/0/11 and GigabitEthernet 1/0/12 are connected to access point (AP) devices.
- The power priority of GigabitEthernet 1/0/2 is critical.
- The power of the AP device connected to GigabitEthernet 1/0/11 does not exceed 9,000 milliwatts.

Network diagram

Figure 355 Network diagram for PoE



Configuration procedure

Enable PoE on GigabitEthernet 1/0/1, GigabitEthernet 1/0/2, GigabitEthernet 1/0/11, and GigabitEthernet 1/0/12.

```
<Sysname> system-view
[Sysname] interface GigabitEthernet 1/0/1
```

```
[Sysname-GigabitEthernet1/0/1] poe enable
[Sysname-GigabitEthernet1/0/1] quit
[Sysname] interface GigabitEthernet 1/0/2
[Sysname-GigabitEthernet1/0/2] poe enable
[Sysname-GigabitEthernet1/0/2] quit
[Sysname] interface GigabitEthernet 1/0/11
[Sysname-GigabitEthernet1/0/11] poe enable
[Sysname-GigabitEthernet1/0/11] quit
[Sysname] interface GigabitEthernet 1/0/12
[Sysname-GigabitEthernet1/0/12] poe enable
[Sysname-GigabitEthernet1/0/12] quit
```

Set the power priority level of GigabitEthernet 1/0/2 to **critical**.

```
<Sysname> system-view
[Sysname] interface GigabitEthernet 1/0/2
[Sysname-GigabitEthernet1/0/2] poe priority critical
[Sysname-GigabitEthernet1/0/2] quit
```

Set the maximum power of GigabitEthernet 1/0/11 to 9,000 milliwatts.

```
[Sysname] interface GigabitEthernet 1/0/11
[Sysname-GigabitEthernet1/0/11] poe max-power 9000
[Sysname-GigabitEthernet1/0/11] quit
```

After the configuration takes effect, the IP phone and AR device are powered and can work normally.

Troubleshooting PoE

Symptom 1: Setting of the priority of a PoE interface to **critical** fails.

Analysis:

- The guaranteed remaining power of the PSE is lower than the maximum power of the PoE interface.
- The priority of the PoE interface is already set.

Solution:

- In the first case, you can solve the problem by reducing the maximum power of the PoE interface when the guaranteed remaining power of the PSE cannot be modified.
- In the second case, you should first remove the priority already configured.

Symptom 2: Applying a PoE configuration file to a PoE interface fails.

Analysis:

- Some configurations in the PoE configuration file are already configured.
- Some configurations in the PoE configuration file do not meet the configuration requirements of the PoE interface.
- Another PoE configuration file is already applied to the PoE interface.

Solution:

- In the first case, you can solve the problem by removing the original configurations of those configurations.
- In the second case, you need to modify some configurations in the PoE configuration file.
- In the third case, you need to remove the application of the undesired PoE configuration file to the PoE interface.

94

sFLOW CONFIGURATION

When configuring sFlow, go to these sections for information you are interested in:

- "sFlow Overview" on page 1203
- "Configuring sFlow" on page 1204
- "Displaying sFlow" on page 1204
- "sFlow Configuration Example" on page 1204
- "Troubleshooting sFlow Configuration" on page 1206

sFlow Overview

Introduction to sFlow

Based on packet sampling, Sampled Flow (sFlow) is a traffic monitoring technology mainly used to collect and analyze traffic statistics.

sFlow has the following two sampling mechanisms:

- Packet-based sampling: Samples one packet out of a specified number of packets from an sFlow enabled port.
- Time-based sampling: Samples interface statistics at a specified interval from an sFlow enabled port.

The sFlow system involves an sFlow agent embedded in a device and a remote sFlow collector. The sFlow agent collects traffic from the sFlow enabled ports, encapsulates the information into sFlow packets, and sends the packets to the sFlow collector. The sFlow collector analyzes the sFlow packets and displays the results.

As a traffic monitoring technology, sFlow has the following advantages:

- Supports traffic monitoring on Gigabit and higher-speed networks.
- Provides scalability with one sFlow collector monitoring multiple or more sFlow agents.
- Implements the low-cost sFlow agent.



Currently, only the sFlow agent function is supported on the Switch 4800G Family.

Operation of sFlow sFlow operates as follows:

- 1 With sFlow enabled, a physical port encapsulates received data into packets and sends them to the sFlow agent.
- 2 The sFlow agent periodically collects interface statistics on all sFlow enabled ports.
- 3 When the sFlow packet buffer overflows or the one-second timer expires, the sFlow agent sends the sFlow packets to the specified sFlow collector.

Configuring sFlow

Follow these steps to configure sFlow:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure an IP address for the sFlow agent	sflow agent ip <i>ip-address</i>	Required Not configured by default.
Specify the IP address and port number of the sFlow collector	sflow collector ip <i>ip-address</i> [port <i>port-num</i>]	Required Not specified by default.
Set the sFlow interval	sflow interval <i>interval-time</i>	Optional 20 seconds by default.
Enter interface view	interface <i>interface-type</i> <i>interface-number</i>	-
Enable sFlow in the inbound or outbound direction	sflow enable { both inbound outbound }	Required Not enabled by default.
Specify the sFlow sampling mode	sflow sampling-mode { determine random }	Optional random by default. Currently, the determine mode is not supported on the Switch 4800G.
Specify the sFlow sampling rate	sflow sampling-rate <i>rate</i>	Optional 200000 by default.

**CAUTION:**

- *The sFlow agent and sFlow collector must not have the same IP address.*
- *Currently, you can specify at most two sFlow collectors on the Switch 4800G Family.*

Displaying sFlow

To do...	Use the command...	Remarks
Display sFlow configuration information	display sflow	Available in any view

sFlow Configuration Example**Network requirements**

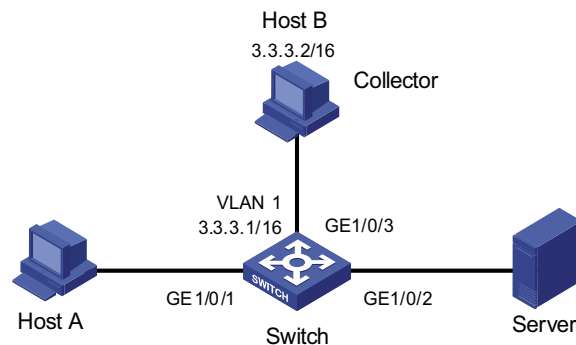
- Host A and Server are connected to Switch through GigabitEthernet 1/0/1 and GigabitEthernet 1/0/2 respectively.
- Host B works as an sFlow collector with IP address 3.3.3.2 and port number 6343, and is connected to Switch through GigabitEthernet 1/0/3.

- GigabitEthernet 1/0/3 belongs to VLAN 1, having an IP address of 3.3.3.1.

Run sFlow agent on Switch, and enable sFlow on GigabitEthernet 1/0/1 to monitor traffic on this interface. Switch sends sFlow packets through GigabitEthernet 1/0/3 to Host B, which then analyzes the sFlow packets and displays the results.

Network diagram

Figure 356 Network diagram for sFlow configuration



Configuration procedure

Configure an IP address for the sFlow agent.

```
<Switch> system-view
[Switch] sflow agent ip 3.3.3.1
```

Specify the IP address and port number of the sFlow collector.

```
[Switch] sflow collector ip 3.3.3.2
```

Set the sFlow interval to 30 seconds.

```
[Switch] sflow interval 30
```

Enable sFlow in both the inbound and outbound directions on GigabitEthernet 1/0/1.

```
[Switch] interface GigabitEthernet 1/0/1
[Switch-GigabitEthernet1/0/1] sflow enable both
```

Specify the traffic sampling rate.

```
[Switch-GigabitEthernet1/0/1] sflow sampling-rate 100000
```

Display the sFlow configuration information.

```
[Switch-GigabitEthernet1/0/1] display sflow
sFlow Global Information:
  Agent          IP:3.3.3.1
  Collector      IP:3.3.3.2      Port: 6343
  Interval(s): 30
sFlow Port Information:
  Interface      Direction      Rate      Mode      Status
  GE1/0/1        Both           100000    Random    Active
```

Troubleshooting sFlow Configuration

The Remote sFlow Collector Cannot Receive sFlow Packets

Symptom

The remote sFlow collector cannot receive sFlow packets.

Analysis

- sFlow is not enabled globally because the sFlow agent or/and the sFlow collector are not specified.
- No port is enabled with sFlow to sample data.
- The IP address of the sFlow collector specified on the sFlow agent is different from that of the remote sFlow collector.
- No IP address is configured for the Layer 3 interface on the device, or the IP address is configured, but the UDP packets with the IP address being the source cannot reach the sFlow collector.
- The physical link between the device and the sFlow collector fails.

Solution

- 1 Check whether sFlow is correctly configured by displaying sFlow configuration with the **display sflow** command.
- 2 Check whether the correct IP address is configured for the device to communicate with the sFlow collector.
- 3 Check whether the physical link between the device and the sFlow collector is normal.

When configuring SSL, go to these sections for information you are interested in:

- “SSL Overview” on page 1207
- “SSL Configuration Task List” on page 1208
- “Displaying and Maintaining SSL” on page 1211
- “Troubleshooting SSL” on page 1211

SSL Overview

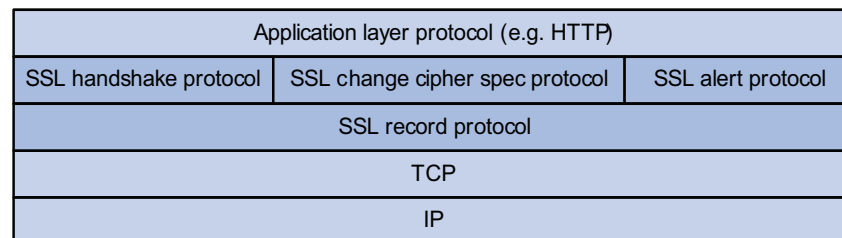
Secure Sockets Layer (SSL) is a security protocol providing secure connection service for TCP-based application layer protocols, for example, HTTP protocol. It is widely used in E-business and online bank fields to provide secure data transmission over the Internet.

SSL provides these security services:

- Confidentiality: SSL encrypts data using a symmetric encryption algorithm and the key generated during the handshake phase.
- Authentication: SSL supports authenticating both the server and the client through certificates, with the authentication of the client being optional.
- Reliability: SSL uses key-based message authentication code (MAC) to verify message integrity.

As shown in Figure 357, the SSL protocol consists of two layers of protocols: the SSL record protocol at the lower layer and the SSL handshake protocol, change cipher spec protocol, and alert protocol at the upper layer.

Figure 357 SSL protocol stack



- SSL handshake protocol: Responsible for establishing a session between a client and the server. A session consists of a set of parameters such as the session ID, peer certificate, cipher suite (including key exchange algorithm, data encryption algorithm and MAC algorithm), compression algorithm, and master key. An SSL session can be used to establish multiple connections, reducing session negotiation cost.

- SSL change cipher spec protocol: Used for notification between a client and the server that the subsequent packets are to be protected and transmitted based on the newly negotiated cipher suite and key.
- SSL alert protocol: Allowing a client and the server to send alert messages to each other. An alert message contains the alert severity level and a description.
- SSL record protocol: Fragmenting and compressing data to be transmitted, calculating and adding MAC to the data, and encrypting the data before transmitting it to the peer end.

SSL Configuration Task List

Different parameters are required on the SSL server and the SSL client.

Complete the following tasks to configure SSL:

Task	Remarks
"Configuring an SSL Server Policy" on page 1208	Required
"Configuring an SSL Client Policy" on page 1210	Optional

Configuring an SSL Server Policy

An SSL server policy is a set of SSL parameters for a server to use when booting up. An SSL server policy takes effect only after it is associated with an application layer protocol, HTTP protocol, for example.

Configuration Prerequisites

Before configuring an SSL server policy, you must configure a PKI (public key infrastructure) domain.

Configuration Procedure

Follow these steps to configure an SSL server policy:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create an SSL server policy and enter its view	ssl server-policy <i>policy-name</i>	Required
Specify a PKI domain for the SSL server policy	pki-domain <i>domain-name</i>	Required By default, no PKI domain is specified for an SSL server policy.
Specify the cipher suite(s) for the SSL server policy to support	ciphersuite [rsa_aes_128_cbc_sha rsa_des_cbc_sha rsa_rc4_128_md5 rsa_rc4_128_sha] *	Optional By default, an SSL server policy supports all cipher suites.
Set the handshake timeout time for the SSL server	handshake timeout <i>time</i>	Optional 3,600 seconds by default
Configure the SSL connection close mode	close-mode wait	Optional Not wait by default

To do...	Use the command...	Remarks
Set the maximum number of cached sessions and the caching timeout time	session { <i>cache-size</i> size <i>timeout</i> time } *	Optional The defaults are as follows: 500 for the maximum number of cached sessions, 3600 seconds for the caching timeout time.
Enable certificate-based SSL client authentication	client-verify enable	Optional Not enabled by default



If you enable client authentication here, you must request a local certificate for the client.

SSL Server Policy Configuration Example

Network requirements

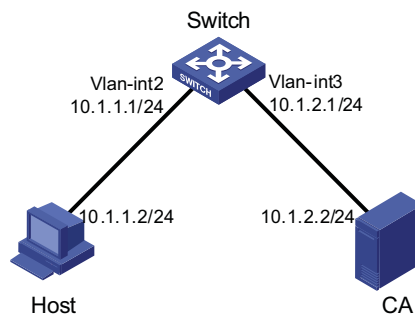
- A switch works as the HTTPS server.
- A host works as the client and accesses the HTTPS server through HTTP secured with SSL.
- A certificate authentication (CA) issues a certificate to the switch.



CAUTION: In this instance, Windows Server works as the CA and the Simple Certificate Enrollment Protocol (SCEP) plug-in is installed on the CA.

Network diagram

Figure 358 Network diagram for SSL server policy configuration



Configuration procedure

- 1 Request a certificate for the switch

Create a PKI entity named **en** and configure it.

```

<Sysname> system-view
[Sysname] pki entity en
[Sysname-pki-entity-en] common-name http-server1
[Sysname-pki-entity-en] fqdn ssl.security.com
[Sysname-pki-entity-en] quit
  
```

Create a PKI domain and configure it.

```

[Sysname] pki domain 1
[Sysname-pki-domain-1] ca identifier ca1
[Sysname-pki-domain-1] certificate request url http://10.1.2.2/certsrv/mscep/mscep.dll
  
```

```
[Sysname-pki-domain-1] certificate request from ra
[Sysname-pki-domain-1] certificate request entity en
[Sysname-pki-domain-1] quit
```

Create a local key pair through RSA.

```
[Sysname] public-key local create rsa
```

Retrieve the CA certificate.

```
[Sysname] pki retrieval-certificate ca domain 1
```

Request a local certificate.

```
[Sysname] pki request-certificate domain 1
```

2 Configure an SSL server policy

Create an SSL server policy named myssl.

```
[Sysname] ssl server-policy myssl
```

Specify the PKI domain for the SSL server policy as 1.

```
[Sysname-ssl-server-policy-myssl] pki-domain 1
```

Enable client authentication.

```
[Sysname-ssl-server-policy-myssl] client-verify enable
[Sysname-ssl-server-policy-myssl] quit
```

3 Associate HTTPS service with the SSL server policy and enable HTTPS service

Configure HTTPS service to use SSL server policy myssl.

```
[Sysname] ip https ssl-server-policy myssl
```

Enable HTTPS service.

```
[Sysname] ip https enable
```

4 Verify your configuration

Launch IE on the host and enter `https://10.1.1.1` in the address bar. You should be able to log in to the switch and manage it.



- For details about PKI configuration commands, refer to “PKI Configuration” on page 1219.
- For details about the **public-key local create rsa** command, refer to “SSH Configuration” on page 1107.

Configuring an SSL Client Policy

An SSL client policy is a set of SSL parameters for a client to use when connecting to the server. An SSL client policy takes effect only after it is associated with an application layer protocol.

Configuration Prerequisites

Before configuring an SSL client policy, you must configure a PKI domain. For details about PKI domain configuration, refer to “Configuring a PKI Domain” on page 1223.

Configuration Procedure

Follow these steps to configure an SSL client policy:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create an SSL client policy and enter its view	ssl client-policy <i>policy-name</i>	Required
Specify a PKI domain for the SSL client policy	pki-domain <i>domain-name</i>	Required No PKI domain is configured by default.
Specify the preferred cipher suite for the SSL client policy	prefer-cipher { rsa_aes_128_cbc_sha rsa_des_cbc_sha rsa_rc4_128_md5 rsa_rc4_128_sha }	Optional rsa_rc4_128_md5 by default
Specify the SSL protocol version for the SSL client policy	version { ssl3.0 tls1.0 }	Optional TLS 1.0 by default



If you enable client authentication on the server, you must request a local certificate for the client.

Displaying and Maintaining SSL

To do...	Use the command...	Remarks
Display SSL server policy information	display ssl server-policy { <i>policy-name</i> all }	Available in any view
Display SSL client policy information	display ssl client-policy { <i>policy-name</i> all }	

Troubleshooting SSL

SSL Handshake Failure Symptom

As the SSL server, the device fails to handshake with the SSL client.

Analysis

SSL handshake failure may result from the following causes:

- No SSL server certificate exists, or the certificate is not trusted.
- The server is expected to authenticate the client, but the SSL client has no certificate or the certificate is not trusted.
- The cipher suites used by the server and the client do not match.

Solution

- 1 You can issue the **debugging ssl** command and view the debugging information to locate the problem:
- 2 If the SSL server has no certificate, request one for it.

- 3 If the server certificate cannot be trusted, install on the SSL client the root certificate of the CA that issues the local certificate to the SSL server, or let the server requests a certificate from the CA that the SSL client trusts.
- 4 If the SSL server is configured to authenticate the client, but the certificate of the SSL client does not exist or cannot be trusted, request and install a certificate for the client.
- 5 You can use the **display ssl server-policy** command to view the cipher suite used by the SSL server policy. If the cipher suite used by the SSL server does not match that used by the client, use the **ciphersuite** command to modify the cipher suite of the SSL server.

96

HTTPS CONFIGURATION

When configuring HTTPS, go to these sections for information you are interested in:

- "HTTPS Overview" on page 1213
- "HTTPS Configuration Task List" on page 1213
- "Associating the HTTPS Service with an SSL Server Policy" on page 1214
- "Enabling the HTTPS Service" on page 1214
- "Associating the HTTPS Service with a Certificate Attribute Access Control Policy" on page 1215
- "Associating the HTTPS Service with an ACL" on page 1215
- "Displaying and Maintaining HTTPS" on page 1215
- "HTTPS Configuration Example" on page 1215

HTTPS Overview

The HTTP Security (HTTPS) refers to the HTTP protocol that supports the Security Socket Layer (SSL) protocol.

The SSL protocol of HTTPS enhances the security of the device in the following ways:

- Uses the SSL protocol to ensure the legal clients to access the device securely and prohibit the illegal clients;
- Encrypts the data exchanged between the HTTPS client and the device to ensure the data security and integrity, thus realizing the security management of the device;
- Defines certificate attribute-based access control policy for the device to control the access right of the client, in order to further avoid attacks from illegal clients.



The total number of HTTP connections and HTTPS connections on a device cannot exceed five.

HTTPS Configuration Task List

Complete these tasks to configure HTTPS:

Configuration task	Remarks
"Associating the HTTPS Service with an SSL Server Policy" on page 1214	Required
"Enabling the HTTPS Service" on page 1214	Required
"Associating the HTTPS Service with a Certificate Attribute Access Control Policy" on page 1215	Optional

Configuration task	Remarks
"Associating the HTTPS Service with an ACL" on page 1215	Optional

Associating the HTTPS Service with an SSL Server Policy

You need to associate the HTTPS service with a created SSL server policy before enabling the HTTPS service.

Follow these steps to associate the HTTPS service with an SSL server policy:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Associate the HTTPS service with an SSL server policy	ip https ssl-server-policy <i>policy-name</i>	Required Not associated by default



- If the **ip https ssl-server-policy** command is executed repeatedly, the HTTPS service is only associated with the last specified SSL server policy.
- When the HTTPS service is disabled, the association between the HTTPS service and the SSL server is automatically removed. To enable it again, you need to re-associate the HTTPS service with an SSL server policy.
- When the HTTPS service is enabled, no modification of its associated SSL server policy takes effect.

Enabling the HTTPS Service

Before configuring the HTTPS, make sure that the HTTPS server is enabled. Otherwise, other related configurations cannot take effect.

Follow these steps to enable the HTTPS service:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enable the HTTPS service	ip https enable	Required Disabled by default.



- After the HTTPS service is enabled, you can use the **display ip https** command to view the state of the HTTPS service and verify the configuration.
- Enabling of the HTTPS service will trigger an SSL handshake negotiation process. During the process, if the local certificate of the device already exists, the SSL negotiation is successfully performed, and the HTTPS service can be started normally. If no local certificate exists, a certificate application process will be triggered by the SSL negotiation. Since the application process takes much time, the SSL negotiation may fail and the HTTPS service cannot be started normally. Therefore, the **ip https enable** command must be executed for multiple times to ensure normal startup of the HTTPS service.

Associating the HTTPS Service with a Certificate Attribute Access Control Policy

Associating the HTTPS service with a configured certificate access control policy helps control the access right of the client, thus providing the device with enhanced security.

Follow these steps to associate the HTTPS service with a certificate attribute access control policy:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Associate the HTTPS service with a certificate attribute access control policy	ip https certificate access-control-policy <i>policy-name</i>	Required Not associated by default.



- If the **ip https certificate access-control-policy** command is executed repeatedly, the HTTPS server is only associated with the last specified certificate attribute access control policy.
- If the HTTPS service is associated with a certificate attribute access control policy, the **client-verify enable** command must be configured in the SSL server policy. Otherwise, the client cannot log onto the device.
- If the HTTPS service is associated with a certificate attribute access control policy, the latter must contain at least one **permit** rule. Otherwise, no HTTPS client can log onto the device.
- For the configuration of an SSL server policy, refer to “PKI Configuration” on page 1219.

Associating the HTTPS Service with an ACL

Associating the HTTPS service with an ACL can filter out requests from some clients to let pass only clients that pass the ACL filtering.

Follow these steps to associate the HTTPS service with an ACL:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Associate the HTTPS service with an ACL	ip https acl <i>acl-number</i>	Required Not associated by default.



If the **ip https acl** command is executed repeatedly, the HTTPS service is only associated with the last specified ACL.

Displaying and Maintaining HTTPS

To do...	Use the command...	Remarks
Display information about HTTPS	display ip https	Available in any view

HTTPS Configuration Example

Network requirements

- Host acts as the HTTPS client and Switch acts as the HTTPS server.

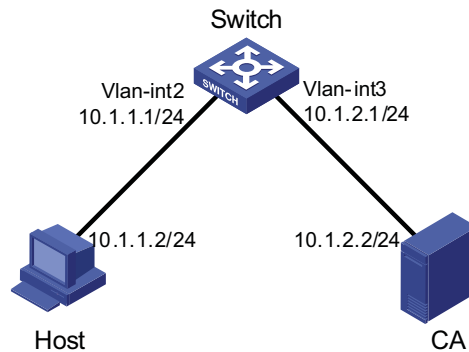
- Host accesses Switch through Web to control Switch.
- CA (Certificate Authority) issues certificate to Switch. The common name of CA is **new-ca**.



CAUTION: In this configuration example, Windows Server serves as CA and you need to install Simple Certificate Enrollment Protocol (SCEP) component.

Network diagram

Figure 359 Network diagram for HTTPS configuration



Configuration procedure

Perform the following configurations on Switch:

1 Apply for a certificate for Switch

Configure a PKI entity.

```

<Switch> system-view
[Switch] pki entity en
[Switch-pki-entity-en] common-name http-server1
[Switch-pki-entity-en] fqdn ssl.security.com
[Switch-pki-entity-en] quit
  
```

Configure a PKI domain.

```

[Switch] pki domain 1
[Switch-pki-domain-1] ca identifier ca1
[Switch-pki-domain-1] certificate request url http://10.1.2.2:8080/certsrv/mscep/mscep.dll
[Switch-pki-domain-1] certificate request from ra
[Switch-pki-domain-1] certificate request entity en
[Switch-pki-domain-1] quit
  
```

Generate a key pair locally by using the RSA algorithm.

```
[Switch] public-key local create rsa
```

Obtain a server certificate from CA.

```
[Switch] pki retrieval-certificate ca domain 1
```

Apply for a local certificate.

```
[Switch] pki request-certificate domain 1
```

2 Configure an SSL server policy associated with the HTTPS service

Configure SSL server policy.

```
[Switch] ssl server-policy myssl
[Switch-ssl-server-policy-myssl] pki-domain 1
[Switch-ssl-server-policy-myssl] client-verify enable
[Switch-ssl-server-policy-myssl] quit
```

3 Configure certificate access control policy

Configure certificate attribute group.

```
[Switch] pki certificate attribute-group mygroup1
[Switch-pki-cert-attribute-group-mygroup1] attribute 1 issuer-name dn ctn new-ca
[Switch-pki-cert-attribute-group-mygroup1] quit
```

Configure certificate access control policy **myacp** and create a control rule.

```
[Switch] pki certificate access-control-policy myacp
[Switch-pki-cert-acp-myacp] rule 1 permit mygroup1
[Switch-pki-cert-acp-myacp] quit
```

4 Reference an SSL server policy

Associate the HTTPS service with the SSL server policy **myssl**.

```
[Switch] ip https ssl-server-policy myssl
```

5 Associate the HTTPS service with a certificate attribute access control policy

Associate the HTTPS service with a certificate attribute access control policy **myacp**.

```
[Switch] ip https certificate access-control-policy myacp
```

6 Enable the HTTPS service

Enable the HTTPS service.

```
[Switch] ip https enable
```

7 Verify the configuration

Launch the IE explorer on Host, and enter https://10.1.1.1. You can log onto Switch and control it.



- For details of PKI commands, refer to PKI "PKI Configuration" on page 1219.
- For details of the **public-key local create rsa** command, refer to "SSH Configuration" on page 1107.

When configuring PKI, go to these sections for information you are interested in:

- “Introduction to PKI” on page 1219
- “PKI Configuration Task List” on page 1222
- “Displaying and Maintaining PKI” on page 1229
- “PKI Configuration Examples” on page 1230
- “Troubleshooting PKI” on page 1235

Introduction to PKI

This section covers these topics:

- “PKI Overview” on page 1219
- “PKI Terms” on page 1219
- “Architecture of PKI” on page 1220
- “Applications of PKI” on page 1221
- “Operation of PKI” on page 1221

PKI Overview

Public Key Infrastructure (PKI) is a system designed for providing information security through public key technologies and digital certificates and verifying the identities of the digital certificate owners.

PKI employs digital certificates, which are bindings of certificate owner identity information and public keys. PKI allows users to request certificates, use certificates, and revoke certificates. By leveraging digital certificates and relevant services like certificate distribution and blacklist publication, PKI supports authentication the entities involved in communication, and thus guaranteeing the confidentiality, integrity and non-repudiation of data.

PKI Terms **Digital certificate**

A digital certificate is a file signed by a certificate authority (CA) that contains a public key and the related user identity information. A simplest digital certificate contains a public key, an entity name, and a digital signature from the CA. Generally, a digital certificate also includes the validity period of the key, the name of the CA and the sequence number of the certificate. A digital certificate must comply with the international standard of ITUTX.5.9. This manual involves two types of certificates: local certificate and CA certificate. A local certificate is a digital certificate signed by a CA for an entity, while a CA certificate, also known as root certificate, is signed by the CA for itself.

CRL

An existing certificate may need to be revoked when, for example, the user name changes, the private key leaks, or the user stops the business. Revoking a certificate is to remove the binding of the public key with the user identity information. In PKI, the revocation is made well known through certificate revocation lists (CRLs). Whenever a certificate is revoked, the CA publishes one or more CRLs to announce that the certificate is invalid. The CRLs contains the serial numbers of all certificates that are revoked and function an effective way for checking the validity of certificates.

A CA may publish multiple CRLs when the number of revoked certificates is so large that publishing them in a single CRL may degrade network performance.

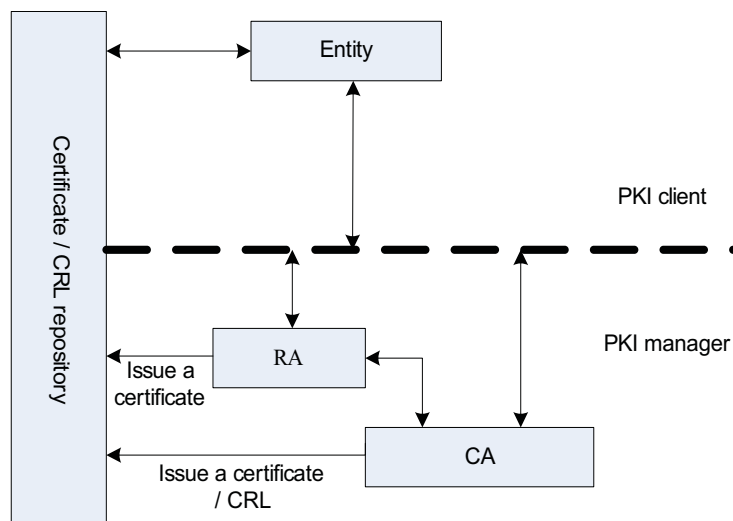
CA policy

A CA policy is a set of criteria that a CA follows in managing certificate requests and in issuing, revoking, and publishing CRLs. Usually, a CA advertises its policy in the form of certification practice statement (CPS), which can be acquired through out-of-band means such as phone, disk, and e-mail or through other means. Since different CAs may use different methods to check the binding of a public key with an entity, make sure that you understand the CA policy before selecting a trusted CA for certificate request.

Architecture of PKI

A PKI system consists of entities, a CA, a registration authority (RA) and a PKI repository, as shown in Figure 360.

Figure 360 PKI architecture



Entity

An entity is an end user of PKI products or services, such as a person, an organization, a device like a switch, or a process running on a computer.

CA

A CA is a trusted entity responsible for issuing and managing digital certificates. A CA issues certificates, specifies the validity period of a certificate, and revokes a certificate as needed by publishing CRLs.

RA

A registration authority (RA) is an extended part of a CA or an independent authority. An RA can implement functions including identity authentication, CRL management, key pair generation and key pair backup. The PKI standard recommends that an independent RA be used for registration management to achieve higher security of application systems.

PKI repository

A PKI repository includes a Lightweight Directory Access Protocol (LDAP) server and some common databases that stores and manages information like certificate requests, certificates, keys, CRLs and logs while providing a simple query function.

LDAP is a protocol for accessing and managing PKI information. An LDAP server stores user information and digital certificates from the RA server and provides directory navigation service. From an LDAP server, an entity can retrieve local and CA certificates of its own as well as certificates of other entities.

Applications of PKI

The PKI technology can satisfy the security requirements of online transactions. As an infrastructure, PKI has a wide range of applications. Here are some application examples.

VPN

A virtual private network (VPN) is a proprietary data communication network built over the public communication infrastructure. A VPN can leverage network layer security protocols (for instance, IPSec) in conjunction with PKI-based encryption and digital signature technologies for confidentiality.

Secure E-mail

E-mails also require confidentiality, integrity, authentication, and non-repudiation. PKI can address these needs. The secure E-mail protocol that is currently developing rapidly is Secure/Multipurpose Internet Mail Extensions (S/MIME), which is based on PKI and allows for transfer of encrypted mails and mails with signature.

Web security

For Web security, two peers can establish a Secure Sockets Layer (SSL) connection first for transparent and secure communications at the application layer. With PKI, SSL enables communications with encryption between a browser and a server. Both the communication parties can identify the identity of each other through digital certificates.

Operation of PKI

In a PKI-enabled network, an entity can request a local certificate from the CA and the device can check the validity of certificates. Here is how it works:

- 1 An entity submits a certificate request to the CA.
- 2 RA reviews the identity of the entity and then sends the identity information and the public key with a digital signature to the CA.
- 3 The CA validates the digital signature, approves the application, and issues a certificate.

- 4 The RA receives the certificate from the CA, sends it to the LDAP server to provide directory navigation service, and notifies the entity that the certificate is successfully issued.
- 5 The entity retrieves the certificate. With the certificate, the entity can communicate with other entities safely through encryption and digital signature.
- 6 The entity makes a request to the CA when it needs to revoke its certificate, while the CA approves the request, updates the CRLs and transfers the CRLs to the LDAP server.

PKI Configuration Task List

Complete the following tasks to configure PKI:

Task	Remarks
"Configuring an Entity DN" on page 1222	Required
"Configuring a PKI Domain" on page 1223	Required
"Submitting a Certificate Request in Auto Mode" on page 1225	"Submitting a Certificate Request in Auto Mode" on page 1225 "Submitting a Certificate Request in Manual Mode" on page 1225
	Required Use either approach
"Retrieving a Certificate Manually" on page 1226	Optional
"Configuring PKI Certificate Validation" on page 1227	Optional
"Destroying a Local RSA Key Pair" on page 1228	Optional
"Deleting a Certificate" on page 1229	Optional
"Configuring an Access Control Policy" on page 1229	Optional

Configuring an Entity DN

A certificate is the binding of a public key and the identity information of an entity, where the identity information is identified by an entity distinguished name (DN). A CA identifies a certificate applicant uniquely by entity DN.

An entity DN is defined by these parameters:

- Common name of the entity.
- Country code of the entity, a standard 2-character code. For example, CN represents China and US represents the United States of America.
- Fully qualified domain name (FQDN) of the entity, a unique identifier of an entity on the network. It consists of a host name and a domain name and can be resolved to an IP address. For example, www.whatever.com is an FQDN, where www is a host name and whatever.com a domain name.
- IP address of the entity.
- Locality where the entity resides.
- Organization to which the entity belongs.
- Unit of the entity in the organization.
- State where the entity resides.



The configuration of an entity DN must comply with the CA certificate issue policy. You need to determine, for example, which entity DN parameters are mandatory and which are optional. Otherwise, certificate request may be rejected.

Follow these steps to configure an entity DN:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create an entity and enter its view	pki entity <i>entity-name</i>	Required No entity exists by default.
Configure the common name for the entity	common-name <i>name</i>	Optional No common name is specified by default.
Configure the country code for the entity	country <i>country-code-str</i>	Optional No country code is specified by default.
Configure the FQDN for the entity	fqdn <i>name-str</i>	Optional No FQDN is specified by default.
Configure the IP address for the entity	ip <i>ip-address</i>	Optional No IP address is specified by default.
Configure the locality of the entity	locality <i>locality-name</i>	Optional No locality is specified by default.
Configure the organization name for the entity	organization <i>org-name</i>	Optional No organization is specified by default.
Configure the unit name for the entity	organization-unit <i>org-unit-name</i>	Optional No unit is specified by default.
Configure the state or province for the entity	state <i>state-name</i>	Optional No state or province is specified by default.



- *Currently, up to two entities can be created on a device.*
- *Windows 2000 CA server has some restrictions on the data length of a certificate request. If the entity DN in a certificate request goes beyond a certain limit, the server does not respond to the certificate request.*

Configuring a PKI Domain

Before requesting a PKI certificate, an entity needs to be configured with some enrollment information, which is referred to as a PKI domain. A PKI domain is intended only for convenience of reference by other applications, and has only local significance.

A PKI domain is defined by these parameters:

- Trusted CA

An entity requests a certificate from a trusted CA.

- Entity

A certificate applicant uses an entity to provide its identity information to a CA.

- RA

Generally, an independent RA is in charge of certificate request management. It receives the registration request from an entity, checks its qualification, and determines whether to ask the CA to sign a digital certificate. The RA only checks the application qualification of an entity; it does not issue any certificate. Sometimes, the registration management function is provided by the CA, in which case no independent RA is required. You are recommended to deploy an independent RA.

- URL of the enrollment server

An entity sends a certificate request to the enrollment server through Simple Certification Enrollment Protocol (SCEP), a dedicated protocol for an entity to communicate with a CA.

- Polling interval and count

After an applicant makes a certificate request, the CA may need a long period of time if it verifies the certificate request manually. During this period, the applicant needs to query the status of the request periodically to get the certificate as soon as possible after the certificate is signed. You can configure the polling interval and count to query the request status.

- IP address of the LDAP server

An LDAP server is usually deployed to store certificates and CRLs. If this is the case, you need to configure the IP address of the LDAP server.

- Fingerprint for root certificate validation

Upon receiving the root certificate of the CA, an entity needs to validate the fingerprint of the root certificate, namely, the hash value of the root certificate content. This hash value is unique to every certificate. The entity will reject the root certificate if the fingerprint of the root certificate does not match the one configured for the PKI domain.

Follow these steps to configure a PKI domain:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a PKI domain and enter its view	pki domain <i>domain-name</i>	Required No PKI domain exists by default.
Specify the trusted CA	ca identifier <i>name</i>	Required No trusted CA is specified by default.
Specify the entity for certificate request	certificate request entity <i>entity-name</i>	Required No entity is specified by default. The specified entity must exist.
Specify the authority for certificate request	certificate request from { ca ra }	Required No authority is specified by default.

To do...	Use the command...	Remarks
Configure the URL of the server for certificate request	certificate request url <i>url-string</i>	Required No URL is configured by default.
Configure the polling interval and maximum number of attempts for querying the certificate request status	certificate request polling { count <i>count</i> interval <i>minutes</i> }	Optional The polling is executed for up to 50 times at the interval of 20 minutes by default.
Specify the LDAP server	ldap-server ip <i>ip-address</i> [port <i>port-number</i>] [version <i>version-number</i>]	Optional No LDP server is specified by default.
Configure the fingerprint for root certificate validation	root-certificate fingerprint { md5 sha1 } <i>string</i>	Optional No fingerprint is configured by default.



- *Currently, up to two PKI domains can be created on a device.*
- *The CA name is required only when you retrieve a CA certificate. It is not used when in local certificate request.*

Submitting a PKI Certificate Request

When requesting a certificate, an entity introduces itself to the CA by providing its identity information and public key, which will be the major components of the certificate that the CA may issue to the entity. A certificate request can be submitted to a CA in two ways: online and offline. In offline mode, a certificate request is submitted to a CA by an "out-of-band" means such as phone, disk, or e-mail.

Online certificate request falls into two categories: manual mode and auto mode.

Submitting a Certificate Request in Auto Mode

In auto mode, an entity automatically requests a certificate through the SCEP protocol when it has no local certificate or the present certificate is about to expire.

Follow these steps to configure an entity to submit a certificate request in auto mode:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PKI domain view	pki domain <i>domain-name</i>	-
Set the certificate request mode to auto	certificate request mode auto [key-length <i>key-length</i> password { cipher simple } <i>password</i>] *	Required Manual by default

Submitting a Certificate Request in Manual Mode

In manual mode, you need to retrieve a CA certificate, generate a local RSA key pair, and submit a local certificate request for an entity.

The goal of retrieving a CA certificate is to verify the authenticity and validity of a local certificate.

Generating an RSA key pair is an important step in certificate request. The key pair includes a public key and a private key. The private key is kept by the user, while the public key is transferred to the CA along with some other information. For detailed information about RSA key pair configuration, refer to “Configuring RSA and DSA Keys” on page 1111.

Follow these steps to submit a certificate request in manual mode:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PKI domain view	pki domain <i>domain-name</i>	-
Set the certificate request mode to manual	certificate request mode manual	Optional Manual by default
Return to system view	quit	-
Retrieve a CA certificate manually	Refer to “Retrieving a Certificate Manually” on page 1226	Required
Generate a local RSA key pair	public-key local create rsa	Required No local RSA key pair exists by default.
Submit a local certificate request	pki request-certificate domain <i>domain-name</i> [<i>password</i>] [pkcs10 [filename <i>filename</i>]]	Required



- If a PKI domain has already a local certificate, creating an RSA key pair will result in inconsistency between the key pair and certificate. To generate a new RSA key pair, delete the local certificate and then issue the **public-key local create rsa** command.
- A newly created key pair will overwrite the existing one. If you perform the **public-key local create rsa** command in the presence of a local RSA key pair, the system will ask you whether you want to overwrite the existing one.
- If a PKI domain has already a local certificate, you cannot request another certificate for it. This is to avoid inconsistency between the certificate and the enrollment information resulting from configuration changes. To request a new certificate, use the **pki delete-certificate** command to delete the existing local certificate and the CA certificate stored locally.
- When it is impossible to request a certificate from the CA through SCEP, you can save the request information by using the **pki request-certificate domain** command with the **pkcs10** and **filename** keywords, and then send the file to the CA by an out-of-band means.
- Make sure the clocks of an entity and the CA are synchronous. Otherwise, the validity period of the certificate may be abnormal.
- The **pki request-certificate domain** configuration will not be saved in the configuration file.

Retrieving a Certificate Manually

You can download an existing CA certificate or local certificate from the CA server and save it locally. To do so, you can use two ways: online and offline. In offline

mode, you need to retrieve a certificate by an out-of-band means like FTP, disk, e-mail and then import it into the local PKI system.

Certificate retrieval serves two purposes:

- Locally store the certificates associated with the local security domain for improved query efficiency and reduced query count;
- Prepare for certificate validation.

Before retrieving a local certificate, be sure to complete LDAP server configuration.

Follow these steps to retrieve a certificate manually:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Retrieve a certificate manually	Online	pki retrieval-certificate { ca local } domain <i>domain-name</i>	Required Use either command
	Offline	pki import-certificate { ca local } domain <i>domain-name</i> { der p12 pem } [filename <i>filename</i>]	



CAUTION:

- If a PKI domain has already a CA certificate, you cannot retrieve another CA certificate for it. This is in order to avoid inconsistency between the certificate and enrollment information due to related configuration changes. To retrieve a new CA certificate, use the **pki delete-certificate** command to delete the existing CA certificate and local certificate first.
- The **pki retrieval-certificate** configuration will not be saved in the configuration file.

Configuring PKI Certificate Validation

A certificate needs to be validated before being used. Validating a certificate is to check that the certificate is signed by the CA and that the certificate has neither expired nor been revoked.

Before validating a certificate, you need to retrieve the CA certificate.

You can specify whether CRL checking is required in certificate validation. If you enable CRL checking, CRLs will be used in validation of a certificate.

Configuring CRL-checking-enabled PKI certificate validation

Follow these steps to configure CRL-checking-enabled PKI certificate validation:

To do...		Use the command...	Remarks
Enter system view		system-view	-
Enter PKI domain view		pki domain <i>domain-name</i>	-

To do...	Use the command...	Remarks
Specify the URL of the CRL distribution point	crl url <i>url-string</i>	Optional No CRL distribution point URL is specified by default.
Set the CRL update period	crl update-period <i>hours</i>	Optional By default, the CRL update period depends on the next update field in the CRL file.
Enable CRL checking	crl check enable	Optional Enabled by default
Return to system view	quit	-
Retrieve the CA certificate	Refer to "Retrieving a Certificate Manually" on page 1226	Required
Retrieve CRLs	pki retrieval-crl domain <i>domain-name</i>	Required
Verify the validity of a certificate	pki validate-certificate { ca local } domain <i>domain-name</i>	Required

Configuring CRL-checking-disabled PKI certificate validation

Follow these steps to configure CRL-checking-disabled PKI certificate validation:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter PKI domain view	pki domain <i>domain-name</i>	-
Disable CRL checking	crl check disable	Required Enabled by default
Return to system view	quit	-
Retrieve the CA certificate	Refer to "Retrieving a Certificate Manually" on page 1226	Required
Verify the validity of the certificate	pki validate-certificate { ca local } domain <i>domain-name</i>	Required



- *The CRL update period refers to the interval at which the entity downloads CRLs from the CRL server. The CRL update period configured manually is prior to that specified in the CRLs.*
- *The **pki retrieval-crl domain** configuration will not be saved in the configuration file.*

Destroying a Local RSA Key Pair

A certificate has a lifetime, which is determined by the CA. When the private key leaks or the certificate is about to expire, you can destroy the old RSA key pair and then create a pair to request a new certificate.

Follow these steps to destroy a local RSA key pair:

To do...	Use the command...	Remarks
Enter system view	system-view	-

To do...	Use the command...	Remarks
Destroy a local RSA key pair	public-key local destroy rsa	Required



For details about the **public-key local destroy rsa** command, refer to “SSH Configuration” on page 1107.

Deleting a Certificate

When a certificate requested manually is about to expire or you want to request a new certificate, you can delete the current local certificate or CA certificate.

Follow these steps to delete a certificate:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Delete certificates	pki delete-certificate { ca local } domain <i>domain-name</i>	Required

Configuring an Access Control Policy

By configuring a certificate attribute-based access control policy, you can further control access to the server, providing additional security for the server.

Follow these steps to configure a certificate attribute-based access control policy:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a certificate attribute group and enter its view	pki certificate attribute-group <i>group-name</i>	Required No certificate attribute group exists by default.
Configure an attribute rule for the certificate issuer name, certificate subject name, or alternative subject name	attribute <i>id</i> { alt-subject-name { fqdn ip } } { issuer-name subject-name } { dn fqdn ip } } { ctn equ nctn nequ } <i>attribute-value</i>	Optional There is no restriction on the issuer name, certificate subject name and alternative subject name by default.
Return to system view	quit	-
Create a certificate attribute-based access control policy and enter its view	pki certificate access-control-policy <i>policy-name</i>	Required No access control policy exists by default.
Configure a certificate attribute-based access control rule	rule [<i>id</i>] { deny permit } <i>group-name</i>	Required No access control rule exists by default.



CAUTION: A certificate attribute group must exist to be associated with a rule.

Displaying and Maintaining PKI

To do...	Use the command...	Remarks
Display the contents or request status of a certificate	display pki certificate { { ca local } domain <i>domain-name</i> request-status }	Available in any view

To do...	Use the command...	Remarks
Display CRLs	display pki crl domain <i>domain-name</i>	Available in any view
Display information about one or all certificate attribute groups	display pki certificate attribute-group { <i>group-name</i> all }	Available in any view
Display information about one or all certificate attribute-based access control policies	display pki certificate access-control-policy { <i>policy-name</i> all }	Available in any view

PKI Configuration Examples



CAUTION:

- The SCEP plug-in is required when you use the Windows Server as the CA. In this case, when configuring the PKI domain, you need to use the **certificate request from ra** command to specify that the entity requests a certificate from an RA.
- The SCEP plug-in is not required when RSA Keon is used. In this case, when configuring a PKI domain, you need to use the **certificate request from ca** command to specify that the entity requests a certificate from a CA.

Configuring a PKI Entity to Request a Certificate from a CA



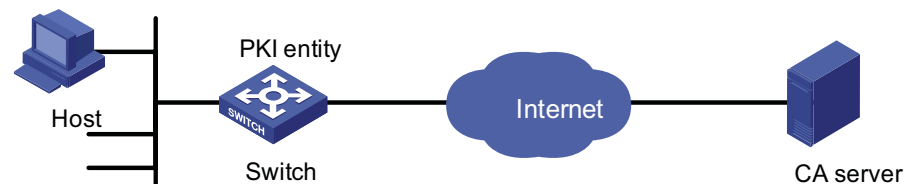
RSA Keon is used on the CA server in this configuration example.

Network requirements

- The device submits a local certificate request to the CA server.
- The device acquires the CRLs for certificate validation.

Network diagram

Figure 361 Diagram for configuring a PKI entity to request a certificate from a CA



Configuration procedure

On the CA server, complete the following configurations:

- 1 Create a CA server named myca

In this example, you need to configure these basic attributes on the CA server at first:

- Nickname: Name of the trusted CA.
- Subject DN: DN information of the CA, including the Common Name (CN), Organization Unit (OU), Organization (O), and Country (C).

The other attributes may be left using the default values.

2 Configure extended attributes

After configuring the basic attributes, you need to perform configuration on the jurisdiction configuration page of the CA server. This includes selecting the proper extension profiles, enabling the SCEP autovetting function, and adding the IP address list for SCEP autovetting.

3 Configure the CRL publishing behavior

After completing the above configuration, you need to perform CRL related configurations. In this example, select the local CRL publishing mode of HTTP and set the HTTP URL to `http://4.4.4.133:447/myca.crl`.

After the above configuration, make sure that the system clock of the device is synchronous to that of the CA, allowing the device to request certificates and retrieve CRLs properly.

On the Switch, perform the following configurations:

4 Configure the entity DN

Configure the entity name as aaa and the common name as Switch.

```
<Switch> system-view
[Switch] pki entity aaa
[Switch-pki-entity-aaa] common-name Switch
[Switch-pki-entity-aaa] quit
```

5 Configure the PKI domain

Create PKI domain torsa and enter its view.

```
[Switch] pki domain torsa
```

Configure the name of the trusted CA as myca.

```
[Switch-pki-domain-torsa] ca identifier myca
```

Configure the URL of the enrollment server in the format of `http://host:port/Issuing Jurisdiction ID`, where Issuing Jurisdiction ID is a hexadecimal string generated on the CA server.

```
[Switch-pki-domain-torsa] certificate request url http://4.4.4.133:446/c95e970f632d27be5e8cbf80e971d9c4a9a93337
```

Set the registration authority to **CA**.

```
[Switch-pki-domain-torsa] certificate request from ca
```

Specify the entity for certificate request as aaa.

```
[Switch-pki-domain-torsa] certificate request entity aaa
```

Configure the URL for the CRL distribution point.

```
[Switch-pki-domain-torsa] crl url http://4.4.4.133:447/myca.crl
[Switch-pki-domain-torsa] quit
```

6 Generate a local key pair using RSA

```
[Switch] public-key local create rsa
The range of public key size is (512 ~ 2048).
NOTES: If the key modulus is greater than 512,
       It may take a few minutes.
Press CTRL+C to abort.
Input the bits in the modulus [default = 1024]:
Generating keys...
.....+++++
.....+++++
.....+++++
.....+++++
.....+++++
.
```

7 Apply for certificates

Retrieve the CA certificate and save it locally.

```
[Switch] pki retrieval-certificate ca domain torsa
Retrieving CA/RA certificates. Please wait a while.....
The trusted CA's finger print is:
    MD5  fingerprint:EDE9 0394 A273 B61A F1B3 0072 A0B1 F9AB
    SHA1 fingerprint: 77F9 A077 2FB8 088C 550B A33C 2410 D354 23B2 73A8

Is the finger print correct?(Y/N):y

Saving CA/RA certificates chain, please wait a moment.....
CA certificates retrieval success.
```

Retrieve CRLs and save them locally.

```
[Switch] pki retrieval-crl domain torsa
Connecting to server for retrieving CRL. Please wait a while.....
CRL retrieval success!
```

Apply for a local certificate manually.

```
[Switch] pki request-certificate domain torsa challenge-word
Enrolling the local certificate,please wait a while.....
Certificate request Successfully!
Saving the local certificate to device.....
Done!
```

8 Verify your configuration

Use the following command to view information about the local certificate acquired.

```
<Switch> display pki certificate local domain torsa
Certificate:
  Data:
    Version: 3 (0x2)
    Serial Number:
```

```

9A96A48F 9A509FD7 05FFF4DF 104AD094
Signature Algorithm: sha1WithRSAEncryption
Issuer:
  C=cn
  O=org
  OU=test
  CN=myca
Validity
  Not Before: Jan  8 09:26:53 2007 GMT
  Not After  : Jan  8 09:26:53 2008 GMT
Subject:
  CN=Switch
Subject Public Key Info:
  Public Key Algorithm: rsaEncryption
  RSA Public Key: (1024 bit)
    Modulus (1024 bit):
      00D67D50 41046F6A 43610335 CA6C4B11
      F8F89138 E4E905BD 43953BA2 623A54C0
      EA3CB6E0 B04649CE C9CDDD38 34015970
      981E96D9 FF4F7B73 A5155649 E583AC61
      D3A5C849 CBDE350D 2A1926B7 0AE5EF5E
      D1D8B08A DBF16205 7C2A4011 05F11094
      73EB0549 A65D9E74 0F2953F2 D4F0042F
      19103439 3D4F9359 88FB59F3 8D4B2F6C
      2B
    Exponent: 65537 (0x10001)
X509v3 extensions:
  X509v3 CRL Distribution Points:
    URI:http://4.4.4.133:447/myca.crl

Signature Algorithm: sha1WithRSAEncryption
836213A4 F2F74C1A 50F4100D B764D6CE
B30C0133 C4363F2F 73454D51 E9F95962
EDE9E590 E7458FA6 765A0D3F C4047BC2
9C391FF0 7383C4DF 9A0CCFA9 231428AF
987B029C C857AD96 E4C92441 9382E798
8FCC1E4A 3E598D81 96476875 E2F86C33
75B51661 B6556C5E 8F546E97 5197734B
C8C29AC7 E427C8E4 B9AAF5AA 80A75B3C

```

You can also use some other **display** commands to view detailed information about the CA certificate and CRLs. Refer to the parts related to **display pki certificate ca domain** and **display pki crl domain** commands in *PKI Commands*.

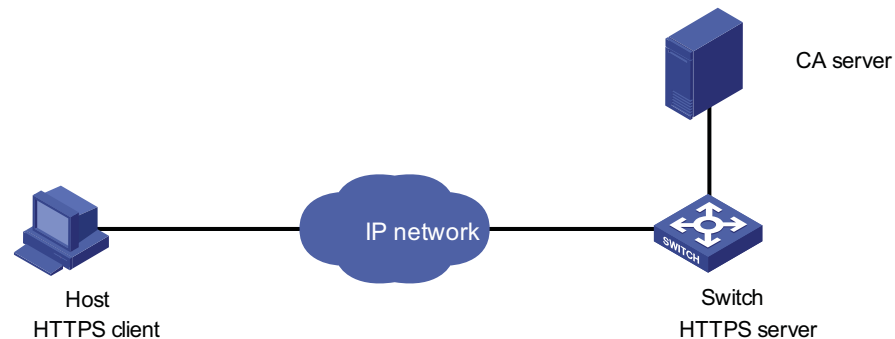
Configuring a Certificate Attribute-Based Access Control Policy

Network requirements

- The client accesses the remote HTTPS server through the HTTP Security (HTTPS) protocol.
- SSL is configured to ensure that only legal clients log into the HTTPS server.
- Create a certificate attribute-based access control policy to control access to the HTTPS server.

Networking diagram

Figure 362 Diagram for configuring a certificate attribute-based access control policy



Configuration procedure



- For detailed information about SSL configuration, refer to “SSL Configuration” on page 1207.
- For detailed information about HTTPS configuration, refer to “HTTPS Configuration” on page 1213.
- The PKI domain to be referenced by the SSL policy must be created in advance. For detailed configuration of the PKI domain, refer to “Configure the PKI domain” on page 1231.
- Configure the HTTPS server

Configure the SSL policy for the HTTPS server to use.

```
<Switch> system-view
[Switch] ssl server-policy myssl
[Switch-ssl-server-policy-myssl] pki-domain 1
[Switch-ssl-server-policy-myssl] client-verify enable
[Switch-ssl-server-policy-myssl] quit
```

1 Configure the certificate attribute group

Create certificate attribute group mygroup1 and add two attribute rules. The first rule defines that the DN of the subject name includes the string aabbcc, and the second rule defines that the IP address of the certificate issuer is 10.0.0.1.

```
[Switch] pki certificate attribute-group mygroup1
[Switch-pki-cert-attribute-group-mygroup1] attribute 1 subject-name
dn ctn aabbcc
[Switch-pki-cert-attribute-group-mygroup1] attribute 2 issuer-name i
p equ 10.0.0.1
[Switch-pki-cert-attribute-group-mygroup1] quit
```

Create certificate attribute group mygroup2 and add two attribute rules. The first rule defines that the FQDN of the alternative subject name does not include the string of apple, and the second rule defines that the DN of the certificate issuer name includes the string aabbcc.

```
[Switch] pki certificate attribute-group mygroup2
[Switch-pki-cert-attribute-group-mygroup2] attribute 1 alt-subject-name fqdn nctn apple
[Switch-pki-cert-attribute-group-mygroup2] attribute 2 issuer-name dn ctn aabbcc
[Switch-pki-cert-attribute-group-mygroup2] quit
```

2 Configure the certificate attribute-based access control policy

Create the certificate attribute-based access control policy of myacp and add two access control rules.

```
[Switch] pki certificate access-control-policy myacp
[Switch-pki-cert-acp-myacp] rule 1 deny mygroup1
[Switch-pki-cert-acp-myacp] rule 2 permit mygroup2
[Switch-pki-cert-acp-myacp] quit
```

3 Apply the SSL server policy and certificate attribute-based access control policy to HTTPS service and enable HTTPS service.

Apply SSL server policy myssl to HTTPS service.

```
[Switch] ip https ssl-server-policy myssl
```

Apply the certificate attribute-based access control policy of myacp to HTTPS service.

```
[Switch] ip https certificate access-control-policy myacp
```

Enable HTTPS service.

```
[Switch] ip https enable
```

Troubleshooting PKI

Failed to Retrieve a CA Certificate

Symptom

Failed to retrieve a CA certificate.

Analysis

Possible reasons include these:

- The network connection is not proper. For example, the network cable may be damaged or loose.
- No trusted CA is specified.
- The URL of the enrollment server for certificate request is not correct or not configured.
- No RA is specified.
- The system clock of the device is not synchronized with that of the CA.

Solution

- Make sure that the network connection is physically proper.
- Check that the required commands are configured properly.
- Use the **ping** command to check that the RA server is reachable.
- Configures the RA for certificate request.
- Synchronize the system clock of the device with that of the CA.

Failed to Request a Local Certificate**Symptom**

Failed to request a local certificate.

Analysis

Possible reasons include these:

- The network connection is not proper. For example, the network cable may be damaged or loose.
- No CA certificate has been retrieved.
- The current key pair has been bound to a certificate.
- No trusted CA is specified.
- The URL of the enrollment server for certificate request is not correct or not configured.
- No RA is configured.
- Some required parameters of the entity DN are not configured.

Solution

- Make sure that the network connection is physically proper.
- Retrieve a CA certificate.
- Regenerate a key pair.
- Specify a trusted CA.
- Use the **ping** command to check that the RA server is reachable.
- Configure the RA for certificate request.
- Configure the required entity DN parameters.

Failed to Retrieve CRLs**Symptom**

Failed to retrieve CRLs.

Analysis

Possible reasons include these:

- The network connection is not proper. For example, the network cable may be damaged or loose.
- No CA certificate has been retrieved before you try to retrieve CRLs.
- The IP address of LDAP server is not configured.
- The URL for CRL distribution is not configured.
- The LDAP server version is wrong.

Solution

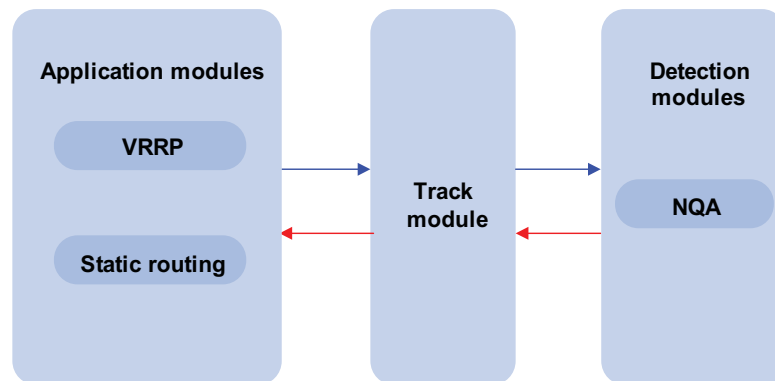
- Make sure that the network connection is physically proper.
- Retrieve a CA certificate.
- Specify the IP address of the LDAP server.
- Specify the URL for CRL distribution.
- Re-configure the LDAP version.

When configuring Track, go to these sections for information you are interested in:

- “Track Overview” on page 1237
- “Track Configuration Task List” on page 1238
- “Configuring Collaboration Between the Track Module and the Detection Modules” on page 1238
- “Configuring Collaboration Between the Track Module and the Application Modules” on page 1239
- “Displaying and Maintaining Track Object(s)” on page 1241
- “Track Configuration Example” on page 1241

Track Overview

Figure 363 Collaboration through the Track module



The Track module is used to implement collaboration between different modules.

The collaboration here involves three parts: the application modules, the Track module, and the detection modules. These modules collaborate with one another through collaboration objects. That is, the detection modules trigger the application modules to perform certain operations through the Track module. More specifically, the detection modules probe the link status, network performance and so on, and inform the application modules of the detection result through the Track module. After the application modules are aware of the changes of network status, they deal with the changes accordingly to avoid communication interruption and network performance degradation.

The Track module works between the application modules and the detection modules and is mainly used to obscure the difference of various detection modules to provide a unified interface for application modules.

Collaboration Between the Track Module and the Detection Modules

You can establish the collaboration between the Track module and the detection modules through configuration. A detection module probes the link status and informs the Track module of the probe result. The Track module then changes the status of the Track object accordingly:

- If the probe succeeds, the status of the corresponding Track object is **Positive**;
- If the probe fails, the status of the corresponding Track object is **Negative**.

At present, the detection modules that can collaborate with the Track module include the Network Quality Analyzer (NQA) only. Refer to “NQA Configuration” on page 1047.

Collaboration Between the Track Module and the Application Modules

You can establish the collaboration between the Track module and the application modules through configuration. If the status of the Track object changes, the Track module tells the application modules to deal with the change accordingly.

At present, the application modules that can collaborate with the Track module include:

- VRRP
- Static routing

Track Configuration Task List

To implement the collaboration function, you need to establish collaboration between the Track module and the detection modules, and between the Track module and the application modules.

Complete these tasks to configure Track module:

Task	Remarks
“Configuring Collaboration Between the Track Module and the Detection Modules” on page 1238	“Configuring Track-NQA Collaboration” on page 1238 Required
“Configuring Collaboration Between the Track Module and the Application Modules” on page 1239	“Configuring Track-VRRP Collaboration” on page 1239 “Configuring Track-Static Routing Collaboration” on page 1240 Use either approach

Configuring Collaboration Between the Track Module and the Detection Modules

Configuring Track-NQA Collaboration

Through the following configuration, you can establish the collaboration between the Track module and the NQA, which probes the link status and informs the Track module of the probe result.

Follow these steps to configure Track-NQA collaboration:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Create a Track object and associate it with the specified Reaction entry of the NQA test group	track track-entry-number nqa entry admin-name operation-tag reaction item-num	Required No Track object is created by default.



CAUTION: When you configure a Track object, the specified NQA test group and Reaction entry can be nonexistent. In this case, the status of the configured Track object is **Invalid**.

Configuring Collaboration Between the Track Module and the Application Modules

Configuring Track-VRRP Collaboration

Through the Track-VRRP collaboration, you can:

- Monitor the upper link. If there is a fault on the upper link of the master of a VRRP group, hosts in the LAN cannot access the external network through the master. In this case, the status of the monitored Track object changes to Negative, and the priority of the master thus decreases by a specified value, allowing a higher priority backup in the VRRP group to become the master to maintain proper communication between the hosts in the LAN and the external network.
- Monitor the master on a backup. If there is a fault on the master, the backup working in the switchover mode will switch to the master immediately to maintain normal communication.

Configuration prerequisites

Before configuring VRRP to monitor a Track object, you need to create a VRRP group on an interface and configure the virtual IP address of the VRRP group.

Configuration procedure

Follow these steps to configure Track-VRRP collaboration:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Enter interface view	interface interface-type interface-number	-
Create a VRRP group and configure its virtual IP address	vrrp vrid virtual-router-id virtual-ip virtual-address	Required No VRRP group is created by default.
Specify a Track object to be monitored by VRRP	vrrp vrid virtual-router-id track track-entry-number [reduced priority-reduced switchover]	Required No Track object is specified for VRRP by default.



- Do not perform Track object monitoring on the IP address owner.
- When the status of the monitored Track object turns from Negative to Positive, the corresponding master restores its priority automatically.
- The monitored Track object can be nonexistent, so that you can first specify the Track object to be monitored using the **vrrp vrid track** command, and then create the Track object using the **track** command.
- Refer to “VRRP Configuration” on page 1073.

Configuring Track-Static Routing Collaboration

You can check the validity of a static route in real time by establishing collaboration between Track and static routing.

If you specify the next hop but not the egress interface when configuring a static route, you can associate the static route with a Track object and thus check the validity of the static route according to the status of the Track object.

- If the status of the Track object is **Positive**, then the next hop of the static route is reachable, and the configured static route is valid.
- If the status of the Track object is **Negative**, then the next hop of the static route is unreachable, and the configured static route is invalid.

Follow these steps to configure the Track-Static Routing collaboration:

To do...	Use the command...	Remarks
Enter system view	system-view	-
Configure the Track-Static Routing collaboration, so as to check the reachability of the next hop of the static route	ip route-static <i>dest-address</i> { <i>mask</i> <i>mask-length</i> } <i>next-hop-address</i> track <i>track-entry-number</i> [preference <i>preference-value</i>] [tag <i>tag-value</i>] [description <i>description-text</i>]	Required Not configured by default.



- For the configuration of Track-Static Routing collaboration, the specified static route can be an existent or nonexistent one. For an existent static route, the static route and the specified Track object are associated directly; for a nonexistent static route, the system creates the static route and then associates it with the specified Track object.
- The Track object to be associated with the static route can be a nonexistent one. After you use the **track** command to create the Track object, the association takes effect.
- If a static route needs route recursion, the associated Track object must monitor the next hop of the recursive route instead of that of the static route; otherwise, a valid route may be considered invalid.
- For details of static route configuration, refer to the “Static Routing Configuration” on page 251.

Displaying and Maintaining Track Object(s)

To do...	Use the command...	Remarks
Display information about the specified Track object or all Track objects	display track { track-entry-number all }	Available in any view

Track Configuration Example

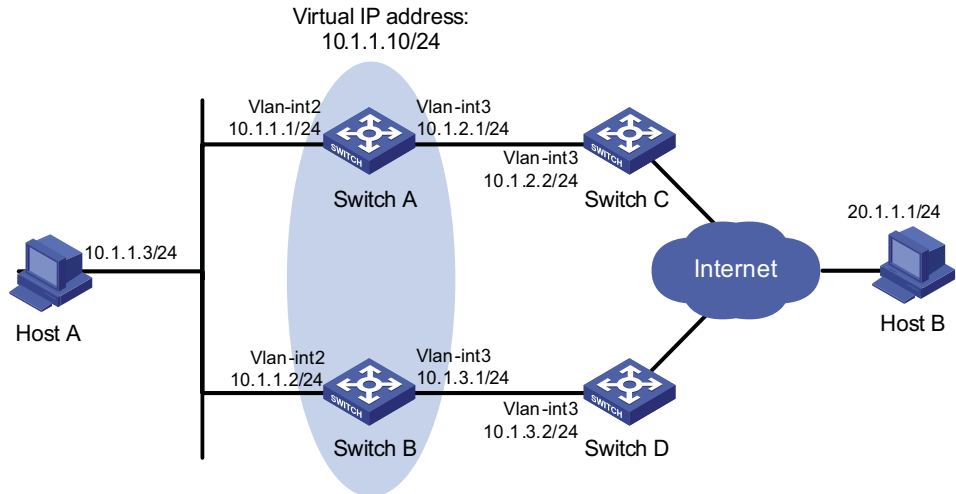
VRRP-Track-NQA Collaboration Configuration Example

Network requirements

- Host A needs to access Host B on the Internet. The default gateway of Host A is 10.1.1.10/24.
- Switch A and Switch B belong to VRRP group 1, whose virtual IP address is 10.1.1.10.
- When Switch A works normally, packets from Host A to Host B are forwarded through Switch A. When VRRP finds that there is a fault on the upper link of Switch A through NQA, packets from Host A to Host B are forwarded through Switch B.

Network diagram

Figure 364 Network diagram for VRRP-Track-NQA collaboration configuration



Configuration procedure

- 1 Configure the IP address of each interface as shown in Figure 364.
- 2 Configure an NQA test group on Switch A.

```
<SwitchA> system-view
```

```
# Create an NQA test group with the administrator name admin and the operation tag test.
```

```
[SwitchA] nqa entry admin test
```

```
# Configure the test type as ICMP-echo.
```

```
[SwitchA-nga-admin-test] type icmp-echo
# Configure the destination address as 10.1.2.2.
[SwitchA-nga-admin-test-icmp-echo] destination ip 10.1.2.2
# Set the test frequency to 100 ms.
[SwitchA-nga-admin-test-icmp-echo] frequency 100
# Configure Reaction entry 1, specifying that five consecutive probe failures
trigger the Track-NQA collaboration.
[SwitchA-nga-admin-test-icmp-echo] reaction 1 checked-element probe-
fail threshold-type consecutive 5 action-type trigger-only
[SwitchA-nga-admin-test-icmp-echo] quit
```

Start NQA probes.

```
[SwitchA] nga schedule admin test start-time now lifetime forever
```

3 Configure a Track object on Switch A.

Configure Track object 1, and associate it with Reaction entry 1 of the NQA test group (with the administrator **admin**, and the operation tag **test**).

```
[SwitchA] track 1 nga entry admin test reaction 1
```

4 Configure VRRP on Switch A.

Create VRRP group 1, and configure the virtual IP address 10.1.1.10 for the group.

```
[SwitchA] interface vlan-interface 2
[SwitchA-Vlan-interface2] vrrp vrid 1 virtual-ip 10.1.1.10
```

Set the priority of Switch A in VRRP group 1 to 110.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 priority 110
```

Set the authentication mode of VRRP group 1 to **simple**, and the authentication key to **hello**.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 authentication-mode simple hello
```

Configure the master to send VRRP packets at an interval of five seconds.

```
[SwitchA-Ethernet1/0] vrrp vrid 1 timer advertise 5
```

Configure Switch A to work in preemptive mode, and set the preemption delay to five seconds.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 preempt-mode timer delay 5
```

Configure to monitor Track object 1 and specify the priority decrement to 30.

```
[SwitchA-Vlan-interface2] vrrp vrid 1 track 1 reduced 30
```

5 Configure VRRP on Switch B.

```

<SwitchB> system-view
[SwitchB] interface vlan-interface 2

# Create VRRP group 1, and configure the virtual IP address 10.1.1.10 for the
group.

[SwitchB-Vlan-interface2] vrrp vrid 1 virtual-ip 10.1.1.10

# Set the authentication mode of VRRP group 1 to simple, and the authentication
key to hello.

[SwitchB-Vlan-interface2] vrrp vrid 1 authentication-mode simple hello

# Configure the master to send VRRP packets at an interval of five seconds.

[SwitchB-Vlan-interface2] vrrp vrid 1 timer advertise 5

# Configure Switch B to work in preemptive mode, and set the preemption delay
to five seconds.

[SwitchB-Vlan-interface2] vrrp vrid 1 preempt-mode timer delay 5

```

6 Verify the configuration

After configuration, ping Host B on Host A, and you can see that Host B is reachable. Use the **display vrrp** command to view the configuration result.

Display detailed information about VRRP group 1 on Switch A.

```

[SwitchA-Vlan-interface2] display vrrp verbose
IPv4 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri      : 110
Preempt Mode    : YES
Auth Type       : SIMPLE TEXT
Track Object    : 1
Virtual IP      : 10.1.1.10
Virtual MAC     : 0000-5e00-0101
Master IP       : 10.1.1.1
Adver. Timer    : 5
State           : Master
Run Pri         : 110
Delay Time      : 5
Key             : hello
Pri Reduced     : 0

```

Display detailed information about VRRP group 1 on Switch B.

```

[SwitchB-Vlan-interface2] display vrrp verbose
IPv4 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri      : 100
Preempt Mode    : YES
Auth Type       : SIMPLE TEXT
Track Object    : 1
Virtual IP      : 10.1.1.10
Virtual MAC     : 0000-5e00-0101
Master IP       : 10.1.1.1
Adver. Timer    : 5
State           : Backup
Run Pri         : 100
Delay Time      : 5
Key             : hello
Pri Reduced     : 0

```

The above output information indicates that in VRRP group 1, Switch A is the master and Switch B is a backup. Packets from Host A to Host B are forwarded through Switch A.

When there is a fault on the link between Switch A and Switch C, you can still successfully ping Host B on Host A. Use the **display vrrp** command to view information about VRRP group 1.

Display detailed information about VRRP group 1 on Switch A when there is a fault on the link between Switch A and Switch C.

```
[SwitchA-Vlan-interface2] display vrrp verbose
IPv4 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri      : 110
Preempt Mode    : YES
Auth Type       : SIMPLE TEXT
Track Object    : 1
Virtual IP      : 10.1.1.10
Master IP       : 10.1.1.2
Adver. Timer    : 5
State           : Backup
Run Pri         : 80
Delay Time      : 5
Key             : hello
Pri Reduced     : 30
```

Display detailed information about VRRP group 1 on Switch B when there is a fault on the link between Switch A and Switch C.

```
[SwitchB-Vlan-interface2] display vrrp verbose
IPv4 Standby Information:
Run Method      : VIRTUAL-MAC
Virtual IP Ping : Enable
Interface       : Vlan-interface2
VRID            : 1
Admin Status    : UP
Config Pri      : 100
Preempt Mode    : YES
Auth Type       : SIMPLE TEXT
Track Object    : 1
Virtual IP      : 10.1.1.10
Virtual MAC     : 0000-5e00-0101
Master IP       : 10.1.1.2
Adver. Timer    : 5
State           : Master
Run Pri         : 100
Delay Time      : 5
Key             : hello
```

The output information indicates that when there is a fault on the link between Switch A and Switch C, the priority of Switch A decreases to 80. Switch A becomes the backup, and Switch B becomes the master. Packets from Host A to Host B are forwarded through Switch B.

A

ACRONYMS

A	
AAA	Authentication, Authorization and Accounting
ABR	Area Border Router
ACL	Access Control List
ARP	Address Resolution Protocol
AS	Autonomous System
ASBR	Autonomous System Border Router
B	
BDR	Backup Designated Router
C	
CAR	Committed Access Rate
CLI	Command Line Interface
CoS	Class of Service
D	
DHCP	Dynamic Host Configuration Protocol
DR	Designated Router
D-V	Distance Vector Routing Algorithm
E	
EGP	Exterior Gateway Protocol
F	
FTP	File Transfer Protocol
G	
GARP	Generic Attribute Registration Protocol
GE	Gigabit Ethernet
GVRP	GARP VLAN Registration Protocol
GMRP	GARP Multicast Registration Protocol
H	
Switch Clustering	3Com Group Management Protocol
I	
IAB	Internet Architecture Board
ICMP	Internet Control Message Protocol
IGMP	Internet Group Management Protocol
IGP	Interior Gateway Protocol
IP	Internet Protocol
L	
LSA	Link State Advertisement

LSDB	Link State DataBase
M	
MAC	Medium Access Control
MIB	Management Information Base
N	
NBMA	Non Broadcast MultiAccess
NIC	Network Information Center
NMS	Network Management System
NVRAM	Nonvolatile RAM
O	
OSPF	Open Shortest Path First
P	
PIM	Protocol Independent Multicast
PIM-DM	Protocol Independent Multicast-Dense Mode
PIM-SM	Protocol Independent Multicast-Sparse Mode
PKI	Public Key Infrastructure
Q	
QoS	Quality of Service
R	
RIP	Routing Information Protocol
RMON	Remote Network Monitoring
RSTP	Rapid Spanning Tree Protocol
S	
SNMP	Simple Network Management Protocol
SP	Strict Priority
SSL	Secure Socket Layer
STP	Spanning Tree Protocol
T	
TCP/IP	Transmission Control Protocol/ Internet Protocol
TFTP	Trivial File Transfer Protocol
ToS	Type of Service
TTL	Time To Live
U	
UDP	User Datagram Protocol
V	
VLAN	Virtual LAN
VOD	Video On Demand
VRRP	Virtual Router Redundancy Protocol
W	
WRR	Weighted Round Robin
X	
XID	eXchange Identification
XRN	eXpandable Resilient Networking
